MGM University

Dr. G. Y. Pathrikar College of Computer Science and

Information Technology, Chhatrapati Sambhajinagar

Final Synopsis

on

"Detection and Classification of Thyroid Cancer using Machine Learning Technique"

Submitted for the award of

Doctor of Philosophy

In

Computer Science and Information Technology

Under the Faculty of Basic and Applied Sciences

Submitted by

Monika Devidas Kate

Under the guidance of

Dr. Vijay Kisanrao Kale
Associate Professor
Dr.G.Y.Pathrikar College of Computer Science and Information Technology,
MGM University**,**Chhatrapati Sambhajinagar

TABLE OF CONTENTS

Sr. No	Topic	Page No
1.	Introduction	3
1.1.	Functioning of Thyroid Gland	3
1.2.	Machine Learning Approaches	5
2.	Problem Statement	7
3.	Motivation of Research	7
4.	Objectives	7
5.	Literature Review	8
5.1	Dataset	8
5.2.	Image Filtration Methods	9-10
5.3.	Feature Extraction Techniques	11-12
5.4.	Machine Learning Techniques (SVM, CNN, ANN, Random Forest, Decision Tree)	12-15
5.5.	Machine Learning Techniques (CNN, MobileNetV2)	15-16
6.	Hypothesis	16
7.	Methodology	17-27
8.	Experiment and Results	27-43
9.	Comparative Result Analysis	44
10.	Conclusion	45
11.	Outline of Thesis	45
12.	References	46-49
13.	Publication	50

1. Introduction

The thyroid gland is the largest endocrine gland in the human body, playing a vital role in regulating metabolism and hormonal balance. Thyroid is also an important public health problem all over the world. Due to this reason, effective diagnosis at an early stage can give the society a better chance of complete survival and healing. Therefore, early thyroid cancer prediction is the key to reduce morbidity and mortality.

1.1 Functioning of Thyroid Gland

The thyroid gland is a small, butterfly-shaped gland located in the neck that plays a vital role in regulating the body's metabolism.

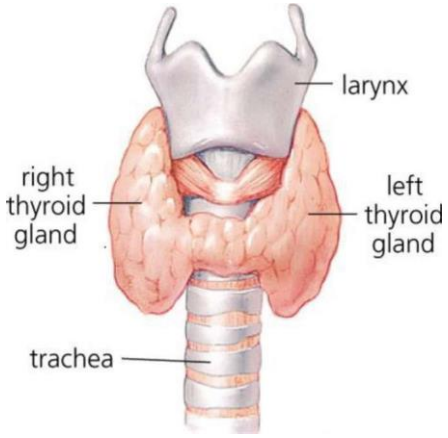


Figure 1 : Thyroid Gland

The thyroid gland is a key endocrine organ that produces thyroxine (T4) and triiodothyronine (T3), hormones essential for regulating metabolism, growth, and energy balance. These hormones influence protein synthesis, oxygen consumption, and thermogenesis, while also playing a role in growth and neural development. Their release is tightly controlled by the hypothalamic–pituitary–thyroid axis through thyroid-stimulating hormone (TSH) feedback regulation (Mullur et al., 2014; Yen, 2001).

1.1.1 Thyroid Disorders

Thyroid disorders occur when the thyroid gland produces either too much or too little hormone, disrupting normal metabolic balance. Hypothyroidism results from insufficient hormone production, leading to fatigue, weight gain, cold intolerance, and slowed metabolism. In contrast, hyperthyroidism is caused by excessive hormone secretion, which accelerates metabolism, resulting in weight loss, heat intolerance, palpitations, and irritability. Other common disorders include goiter (enlargement of the thyroid), thyroid nodules, and thyroid cancer, which may impair hormone regulation and overall endocrine health. (Brent, 2008; Vanderpump, 2011)

1.1.2 Thyroid Cancer

Thyroid cancer is a malignant tumor originating from the cells of the thyroid gland, most commonly from the follicular or parafollicular (C) cells. It is the most frequent endocrine malignancy, accounting for about 3–4% of all newly diagnosed cancers worldwide. The major types include papillary thyroid carcinoma (PTC), follicular thyroid carcinoma (FTC), medullary thyroid carcinoma (MTC), and anaplastic thyroid carcinoma (ATC), with PTC being the most common and ATC the most aggressive. Risk factors include genetic mutations, radiation exposure, family history, and iodine intake imbalance. While most thyroid cancers (especially papillary and follicular) are highly treatable with excellent survival rates, aggressive subtypes such as anaplastic thyroid cancer have poor prognosis (Cabanillas et al., 2016; Sung et al., 2021).

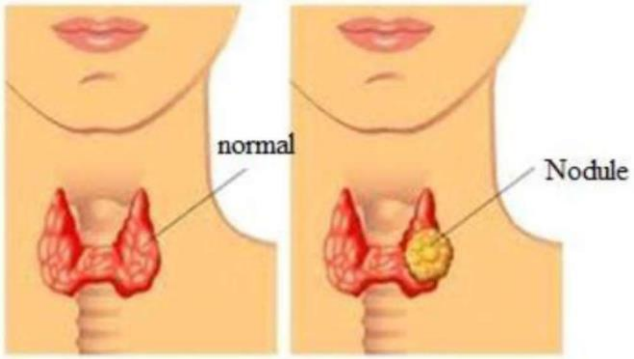


Figure 2: Normal and Abnormal Thyroid Gland

Thyroid cancer is one of the fastest-growing cancers globally, especially among women. According to the World Health Organization and various national cancer registries, thyroid cancer accounts for about 2–3% of all cancers worldwide. In many countries, the incidence rate has been steadily increasing due to better detection methods and increased awareness. However, this rise also reflects overdiagnosis in some cases, making accurate and early detection even more essential.

Early detection of thyroid cancer significantly improves the chances of successful treatment and long-term survival. Most thyroid cancers are highly treatable when identified early, often requiring only surgery and minimal follow-up treatment. However, delayed diagnosis can lead to metastasis (spread to other organs), requiring more aggressive interventions and decreasing the survival rate. Early detection also helps avoid unnecessary procedures in benign cases, reducing the physical and psychological burden on patients.

1.1.3 Thyroid Nodules Types

Thyroid nodules are common, and their prevalence increases with age, although only about 10–15% are malignant. While ultrasound has been a key tool in evaluating nodules for the past two decades, features such as microcalcifications, hypoechogenicity, absence of a halo, and irregular margins are associated with higher malignancy risk. However, interpretation suffers from moderately low inter-rater reliability, which limits consistency.
For cytological evaluation, the Bethesda System for Reporting Thyroid Cytopathology (BSRTC) classifies fine-needle aspiration (FNA) results into five categories, helping to distinguish between benign and malignant nodules for clinical decision-making.

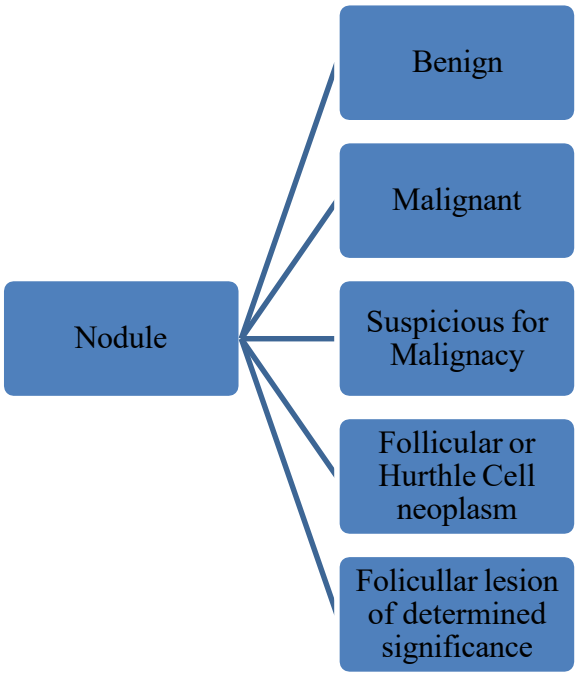


Figure 3: Nodule classification based on Bethesda System

The last mentioned three cytologic diagnosis, accounting for 25percent of all cases, were historically known as indeterminate lesions. They have a 50 percent to 75 percent chance of getting cancer, a 20 percent to 30 percent chance of getting cancer, and a 5 percent to 10% chance of getting cancer, respectively. Colloid nodule, lymphocytic thyroiditis and macro follicular adenoma are the most common benign lesions [7].

1.2 Machine Learning Approaches

In recent years, machine learning has emerged as a promising tool in the medical field, offering the ability to analyze complex patterns in large datasets with high accuracy. This project aims to develop a machine learning model that can effectively detect thyroid cancer using patient data, medical imaging, or relevant biomarkers. By automating and enhancing the diagnostic process, this approach seeks to support medical professionals, reduce human error, and improve early detection rates.

The integration of machine learning in thyroid cancer diagnosis could significantly impact the healthcare system by providing a faster, more reliable, and scalable solution. This research contributes to the ongoing efforts to utilize artificial intelligence in life-saving applications and improve patient care through technological advancement.

Following Point explains how Machine Learning helps to detect thyroid cancer

1. Data Analysis and Pattern Recognition

ML algorithms can process and analyse large and complex medical datasets—such as ultrasound images, cytology reports, and patient clinical features—to identify subtle patterns that may not be easily visible to human experts. These patterns can help differentiate between benign and malignant thyroid nodules with high accuracy.

2. Classification of Thyroid Nodules

ML models are often trained to classify thyroid nodules into categories like benign or malignant using labelled data. For example: Supervised learning algorithms like Support Vector Machines (SVM), Decision Trees, or Neural Networks can be used to classify thyroid tumours based on features like size, shape, texture, and lab values (e.g., TSH, T3, T4 levels).

3. Medical Imaging Interpretation

Using deep learning (especially Convolutional Neural Networks), ML can be applied to interpret ultrasound images of the thyroid gland. These models can detect abnormalities in tissue structure and automatically flag suspicious nodules for further investigation, reducing the workload on radiologists.

4. Reducing Diagnostic Errors

Traditional methods like Fine Needle Aspiration Biopsy (FNAB) can result in indeterminate or inconclusive results. ML can act as a second opinion, improving decision-making by minimizing false positives and false negatives, and suggesting further tests only when necessary.

5. Prediction of Cancer Risk

ML models can also be used to predict the likelihood of thyroid cancer in a patient based on historical data and risk factors (age, gender, family history, hormonal profiles, etc.). This risk scoring helps prioritize patients for further testing or monitoring.

6. Speed and Efficiency

ML algorithms can analyze thousands of cases quickly and consistently, allowing for real-time diagnostic support. This is particularly useful in high-volume clinics or remote areas where access to specialists is limited.

1.2.1 Machine Learning Techniques

Machine learning (ML) is a subset of artificial intelligence that enables systems to learn from data and improve their performance over time without being explicitly programmed. The techniques in ML can be broadly categorized into supervised, unsupervised, semi-supervised, reinforcement, and deep learning approaches. Supervised learning uses labeled data to predict outcomes, with algorithms such as linear regression, decision trees, and support vector machines commonly employed for tasks like classification and regression. Unsupervised learning works with unlabeled data to discover hidden patterns or groupings, using methods like clustering (k-Means, DBSCAN) and dimensionality reduction (PCA, t-SNE). Semi-supervised learning combines a small amount of labeled data with large amounts of unlabeled data to improve model accuracy. Reinforcement learning enables models to learn optimal actions by interacting with an environment and receiving feedback through rewards or penalties, as seen in Q-Learning and policy gradient methods. Deep learning, a specialized subset of ML, employs multi-layered neural networks such as convolutional neural networks (CNNs) for image analysis, recurrent neural networks (RNNs) for sequential data, and transformers for natural language processing. One notable CNN architecture is MobileNetV2, which is designed to be lightweight and efficient, making it particularly

suitable for real-time applications on mobile and embedded devices. MobileNetV2 uses inverted residuals and linear bottlenecks to reduce computation while maintaining high accuracy, making it effective for tasks such as medical image analysis, including thyroid cancer detection. Additionally, ensemble methods like random forests and gradient boosting combine multiple models to enhance predictive performance. Machine learning techniques have diverse applications, ranging from medical diagnosis and fraud detection to recommendation systems and autonomous vehicles, making them central to modern data-driven technologies.

2. Problem Statement

Detecting thyroid cancer traditionally relies on physical exams, ultrasound scans, and fine-needle aspiration biopsies. While these methods are widely used, they can be time-consuming, sometimes invasive, and aren't always accurate—leading to inconclusive or misleading results. As a result, there's a growing need for smarter, more reliable tools that can assist doctors in making faster and more accurate diagnoses. With the rise in available medical data and imaging, there's a real opportunity to develop automated systems that not only improve early detection but also reduce healthcare costs and support better decision-making for both patients and clinicians.

Machine learning can improve thyroid cancer detection by analyzing ultrasound images more accurately and objectively. By identifying key patterns, it helps doctors detect cancer earlier, avoid unnecessary biopsies, and make more reliable decisions—leading to better outcomes for patients.

3. Motivation of Research

Thyroid cancer diagnosis today often relies on methods that can be invasive, slow, and sometimes unreliable, especially where specialists are scarce. This research is motivated by the need to use Computer Aided System to create smarter, faster, and more accurate diagnosis. By helping doctors detect cancer earlier and reduce unnecessary procedures, the goal is to improve patient care and make quality healthcare more accessible.

4. Objectives

- Early prediction of thyroid cancer so that it can help healthcare sector to better assist in treatment of patient.
- To collect and study a comprehensive ultrasound image database of thyroid nodules annotated with severity levels (benign, malignant, etc.) for model training.
- Using the Image processing techniques, classify extracted features into mild, moderate and severe class.
- To classify the malignant and non-malignant thyroid nodules in ultrasound images.

5. Literature Review

5.1 Dataset

Sr. No	Research Paper Title	Author	Type of Dataset used for research	Dataset Details
1	Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction with Machine Learning Approaches	Prabal Poudel	Ultrasound Images	A two 2D thyroid US datasets were used in this work. Dataset 1 consisted of a total of 675 thyroid images with an image size of 760 x 500 pixels were used. This dataset was acquired by a medical expert in SurgicEye GmbH. The second dataset consisted of total of 3370 thyroid US images with an image size of 760 X 1020 pixels. The second dataset was obtained by a thyroid specialist medical doctor at University Clinic of Magdeburg, Germany.
2	Thyroid Disease Prediction Using Machine Learning Approaches	Gyanendra Chaubey	In this research; the ''new-thyroid'' data set is taken which contains 5 attributes and 215 instances. Only two most relevant attributes; total serum thyroxin (T4) and total serum triiodothyronine (T3).	The data set has been taken from the Graven Institute in Sydney, Australia, uploaded to the UC-Irvin. http://archive.ics.uci.edu/ml/machine-learning-databases/thyroiddisease/
3	Cancer Detection using Machine Learning	Pushkar Sathe	Dataset of histopathological 1728 images	https://www.kaggle.com/paultimothymooney/breast-histopathology-images
4	Machine Learning in Healthcare	K. Shailaja		WISCONSIN dataset UCI machine learning repository
5	Thyroid classification and segmentation in ultrasound images using machine learning algorithms	D.Selvathi	The Ultrosound images used in this work are obtained from the online website and some images were obtained from the radiologists. The dataset consists of totally 50 images. 30 thyroid nodule images and 20 normal thyroid images.	Online Repository
6	Interactive Thyroid Disease Prediction System Using Machine Learning Technique	Ankita Tyagi		The data sets for the thyroid diseases have been obtained from the UCI machine learning repository

7	Systematic Poisoning Attacks on and Defenses for Machine Learning in Healthcare	Mehran Mozaffari -Kermani		Five different medical datasets were used from UCI's machine learning repository.
8	Decision Support System for Classification Medullary Thyroid Cancer	Jamil Ahmed Chandio	A real-world dataset was received from Shaheed Muhtrma Benazir Bhutto Medical University, Pakistan(2018-2020). The dataset was comprising over 100 histopathological images of medullary thyroid cancer. Average size of each image was received with x100 magnication power and 4000 X 2500 pixel size per image.	A real-world dataset was received from Shaheed Muhtrma Benazir Bhutto Medical University
9	Cancer Recurrence Prediction using Machine Learing	Shoon Lei Win		Muhtrma Benazir Bhutto Medical University
10	A Comparative study of machine learning algorithms on thyroid disease prediction	Shaik Razia		Disease dataset gathered from the UCI machine learning database was used. http://sci2s.ugr.es/keel/dataset.php?cod=67

5.2 Image Filtration Methods

Sathesh and Rasitha K [35], introduces a nonlinear adaptive median filtering algorithm designed for noise removal in natural images, particularly effective against salt-and-pepper noise. It highlights various denoising techniques and compares their performance based on Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR). The adaptive median filter outperforms other methods by preserving important image details while removing noise, although it is computationally intensive. Experimental results validate its superiority in both visual quality and quantitative measures.

Jae-Chern Yoo and Chang Wook Ahn [36] introduce a "blind-Wiener filter" for image restoration, which operates without prior knowledge of the noise and image power spectra, unlike traditional Wiener filters. The method involves averaging multiple Wiener-filtered versions of a corrupted image, each with added random noise, to enhance signal-to-noise ratio (SNR). The approach demonstrates effectiveness in reducing noise and blur, achieving near-optimal restoration performance comparable to traditional Wiener filters while significantly simplifying the estimation process. Experimental results validate its superiority over existing methods, though the process is computationally intensive.

Arin H. Hamad et. al [37] Denoising of medical images by using some filters. This research work evaluates the effectiveness of various filters—Average, Gaussian, Log, Median, and Wiener—for de-noising medical images affected by Poisson, Speckle, and Gaussian noise. Using MATLAB, the filters were applied to two types of images (cell and breast) in JPG and TIF formats. Results were assessed using image quality metrics like MSE, SNR, and PSNR. The Gaussian filter consistently outperformed others in noise removal,

achieving the highest PSNR values. The research highlights the importance of filter selection based on noise type and confirms the Gaussian filter's superior ability to restore medical images.

Prashant Dwivedi et al. [38] evaluate different filters to remove four types of noise—Salt and Pepper, Gaussian, Rayleigh, and Uniform—from digital images. Each filter's performance is measured using metrics like Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR). Key findings: the Median filter is best for Salt and Pepper noise, the Midpoint filter for Gaussian noise, the Arithmetic Mean filter for Rayleigh noise, and the Geometric Mean filter for Uniform noise. This study helps in selecting the right filter to enhance image quality.

Suman Shrestha [39] proposes a novel adaptive median filter to remove impulse noise (salt-and-pepper noise) from images. Traditional methods often blur the image or struggle with high noise densities. The new method adapts the window size dynamically for better performance. Key highlights are the proposed filter excels in preserving image details while reducing noise. It outperforms standard and adaptive median filters, especially for higher noise densities.

Hetvi Soni et. al [40] The paper compares standard and adaptive median filters for removing image noise. Adaptive median filters outperform standard filters by dynamically adjusting the filter size, making them effective for high noise levels (above 20%) without losing image details. They are particularly good at handling salt-and-pepper noise, though further improvements are needed for Gaussian noise.

Sheikh Tania et.al [41] compares various image filtering techniques (Mean, Median, Wiener filters, Wavelet, and Curvelet transforms) for reducing noise in aerial images. Each method is evaluated based on its ability to remove noise types like Gaussian, salt and pepper, and speckle noise, using metrics like PSNR and MSE. Key findings show that the Median Filter is best for salt and pepper noise, Wavelet Transform works well for speckle noise, and no single method is universally superior. The study highlights the importance of tailored approaches for effective noise removal.

Hanung Adi Nugroho [42] evaluates seven filtering techniques for reducing speckle noise in breast ultrasound images to improve diagnostic accuracy. The m-homog filter showed the best overall performance, ADMSS was effective for edge detection, and DPAD enhanced contrast for lesion detection. Combining filters is suggested for optimal results. This study supports better breast cancer diagnosis through improved image clarity.

Priscila B. Calíope [43] compares various filters for reducing speckle noise in ultrasound images to enhance image quality. It evaluates techniques like the modified MAP filter, Frost filter, and wavelet-based methods. The Frost filter showed strong performance, while the modified MAP filter offered adaptability based on local homogeneity. These findings aim to improve medical ultrasound imaging for better diagnosis.

Shyh-Kuang Ueng [44] presents a structure-based filtering method to enhance ultrasound images by reducing speckle noise and preserving critical features. The approach involves computing the diffusion tensor at each pixel, classifying the local structures, and applying adaptive filters based on the structure type (e.g., 1D Gaussian for edges, 2D Gaussian for spots). Results show improved noise reduction and edge enhancement compared to traditional methods, with iterative passes refining the output for better quality.

S Pradeep [45] provides a review of various speckle noise reduction techniques used in ultrasound medical imaging. These methods are grouped into three main categories: spatial domain techniques, transform domain techniques, and methods based on convolutional neural networks (CNNs). The review discusses the advantages and limitations of each approach and compares their effectiveness in preserving image quality while reducing noise. Additionally, the paper evaluates metrics like Peak Signal-to-Noise Ratio (PSNR) and Mean Squared Error (MSE) for these methods to measure their performance.

5.3 Feature Extraction Techniques

Ajay Kumara[46] This research uses data from the Wisconsin breast cancer dataset available on Kaggle.com to show that utilizing fewer effective features can enhance the accuracy of cancer prediction. This paper involves four steps in total.Total 569 samples are collected in .csv format total 32 features were extracted from samples.In second step data cleansing was performed in which redundantdata is to be removed and missing values are replaced by 0. In third step feature selection process was performed out of 32 features 30 where selected. In next step prediction of cancer was done. In experiment the system specifications are :Python 3.x, Windows 64-bit, 2 GB NVIDIA GeForce graphics card, and Jupiter notebook. All thirty features combine to form the heatmap matrix, sometimes called the correlation matrix.Both the following 11 features and all 30 features are used in the analysis of the data in order to make predictions. The variables that are measured include compactness, concavity, smoothness, perimeter, area, concavity,compactness, and worst of all three: concavity, compactness, and concave points. The data is divided into training and testing portions for analysis, with 80% and 20% going to each group. The dataset of 30 features and 11 features, respectively, is subjected to Support Vector Machine.Based on this research, it can be concluded that selecting the optimal features is an essential component of data analysis to enable more accurate prediction.

Ginni Garg[47] Proposed a method in which 2556 brain MRI images where collected and pre-processed. Segmentation is performed by using Otsu's Thresholding method.15% is used for testing, while 85% of the data is used for training. The feature extraction process involved employing the Gray Level Co-occurrence Matrix (GLCM), Stationary Wavelet Transform (SWT), and Principal Component Analysis (PCA) techniques, which together produced a set of 13 unique features. These features were subsequently utilized in a hybrid ensemble classifier that combines KNN,DT, and Random Forest (RF),classifiers for classification.

Asuntha[48] presented an approach to detect bone cancer on MRI images. Here, the images are smoothed and noise is eliminated using the Gabor filter. Canny edge detector is used for detecting an edge of an image. Morphological operations are essential for recognizing shapes, sizes, and how objects connect in images. Two core operations in morphological processing are dilation, which expands regions, and erosion, which contracts them.This methodology used superpixel segmentation and multilevel segmentation.

This method segments the image into bigger pixels compare to other Title Suppressed Due to Excessive Length 3 segmentation techniques. Various features were extracted from images such as Skewness, Correlation , Root mean square (RMS), Mean , Variance , Inverse difference moment (IDM) ,Standard deviation , Energy , Contrast ,Entropy , Kurtosis , Homogeneity , Smoothness. M. Kunaver[1]This paper presents an overview of various techniques for feature extraction like Spectral features, Shape Features, Texture features. Overview of some advance feature extraction technique is also given like Colour histograms, Fourier coefficients, Morphologi features,Prin- cipal component analysis, Maximum curve points, Maximum curve points,Wavelet functions, Projections.

Ashish Sharma[49]Have work on bone cancer detection using machine learning technique. Proposed work used X-ray images.Median filter is used to remove noise from image. The Canny edge detection technique was employed for segmentation. GLCM technique was used for Feature extraction.This study aimed to extract four key features from the image likecontrast, cor- relation, energy and homogeneity. For bone cancer detection SVM and random forest classifier were used. With 5-fold cross-validation, the model's performance is assessed. Accuracy of SVM is better.

Hitesh Pathak [50] The paper presents a system for detecting lung cancer using CT scan and MRI images. It uses various image processing techniques like preprocessing, segmentation, and feature extraction to improve the output image. Preprocessing includes filters like wiener, median gaussian, Gabor and anisotropic. Segmentation includes Kmean clustering, thresholding, and Fuzzy C-means. Feature extraction includes area, perimeter, and length of major and minor axis of the detected cancer nodule.

Shweta Suresh Naik [51] proposed a system which detect breast cancer using microscopic biopsy images. To detect breast cancer noise removed from biopsy image, next image was segmented using thresholding. In Feature extraction grey level texture features, colour-based features, colour grey level texture features, Law's Texture Energy (LTE) based features, Tamura's features, and wavelet features were extracted. To train and test breast cancer biopsy images machine learning approach is used.
M. Krishna Monika [52]proposed a method to detect skin cancer using machine learning. In this work 800 high- resolution dermoscopic images were collected. In very first step all images were made gray scale image and Gaussian and median filter were employed for noise reduction and image enhancement. Dull Razor Method were used to remove unwanted hair from the skin abnormality. Next, image segmentation is performed using color based k means clustering. GLCM Features were extracted for further classification purposes. MSVM is used for classification. In this work 8 types of skin cancer used for classification. 96.25% accuracy is achieved.

5.4 Machine Learning Techniques

Prabal Poudel et al. [1: have compared the three different machine learning techniques (SVM, ANN and RFC) for thyroid texture classification and seg mentation. In this research two different dataset were used. Dataset 1 consists of total 675 2D thyroid ultrasound images with image size of 760X500 pixel. The Second dataset has 16 subjects and each subject has 156 to 289 2D thyroid US images. For Classification and segmentation of these 2D thyroid US images SVM (Support Vector Machine), ANN (Artificial Neural Network) and RFC (Random Forest classifier algorithms are used. By comparing the performance on same dataset and then on different dataset, comparison analysis was per formed. It is observed that accuracy training of the classifier is similar, ANN perform comparatively better that SVM and RFC. Gyanendra Chaubey et al. [2] have compared widely used three algorithms namely logistic regression, decision trees and k nearest neighbor (kNN) algo rithms to predict thyroid disease and evaluate their performance in terms of accuracy. In this study it is observed that KNN classifier has 96.875% accuracy. Pushkar Sathe et al. [3] have classify tumour into malignant or benign tumour using different features from several cell images. They have fed the machine with categorized data sets of malignant and benign cell images and for better results and have also shuffled these images within themselves and then have successfully trained the machine with malignant and benign data sets after which successful testing is performed with an unknown data set.
K. Shailaja et al. [4] have reviewed various machine learning algorithms for prediction of various diseases. These algorithms are used for developing decision support for healthcare sector.
D. Selvathi et al. [5] This study is focused on to develop an automatic system that classify thyroid images and segments the thyroid gland using machine learning algorithm. Thyroid ultrasound images with nodules and without nod ules dataset was used. In this method mixed thyroid images (with normal thy roid region and thyroid nodules) Were classified using SVM (Support Vector Machine) and Extreme Learning Machine (ELM). During this experiment, the ELM segmentation method is observed with better accuracy than SVM segmentation method.
Ankita Tyagi et al. [6] This research is about to predict thyroid disease using machine learning algorithm like SVM, K-NN, and decision tree. It is observed that Neural Network perform better over other techniques. The performance of Support vector machine and decision trees techniques are also observed good.Vijay Vyas et al. [7] The aim of this research was to develop a computer aided system to classify thyroid nodule as benign or malignant. Thyroid ultrasound 282 M. D. Kate and V. Kale images of 99 patient were used in this research. The SVM and ANN classification algorithms were used as well as both

algorithms were compared based on their accuracy score. From the experimental result it is observed that SVM outperformed ANN, SVM achieved 96% accuracy.

Jamil Ahmed Chandio et al. [8] This paper addresses the classification problem of Medullary Thyroid Cancer and classifies malignant and non-malignant cells and tissues. The proposed methodology of this paper is based upon three layers. Layer 1: Image preprocessing, Layer 2: Classification Model, Layer 3: Result Visualization.

Shoon Lei Win et al. [9] This research study is about to predict cancer recurrence from microarray gene expression data. Here machine learning based approach is used. Gene expression data set used as input to machine learning algorithm. The model will predict whether a specific cancer may reappear within a specific time-frame. In this study three cancer recurrence data set of CNS cancer recurrence, prostate cancer recurrence and breast cancer recurrence were used.

Shaik Razia et al. [10] In this research various Machine learning algorithms like Support Vector Machine (SVM), Multiple Linear Regression, Na¨ ıve Bayes and Decision Trees are used to diagnose thyroid disease. This comparative study results are compared and found that Decision Trees algorithm provided greater accuracy to diagnose thyroid disease.

Sonali Bhadoria et al. [11] This research Proposed an automatic segmen tation method using two tools, analyze 10.0 and mazda for segmentation of thyroid US images and segmentation by applying specific algorithm. As well as comparative analysis is also done. They also provided a summary of all the results.

Reema Mathew A et al. [12] In this review paper comparative study of different feature extraction technique is done for detection of brain tumor. From study it is observed that there are several hybrid approaches can be developed which may result in higher accuracy and better results.

Eystratios G. et al. [13] This study is about to detect thyroid nodule using ultrasound images and videos. Author proposed a computer aided system, named TDN (Thyroid nodule detector). In this research SVM and KNN classification techniques were used, experimental result shows SVM classifier perform better than KNN.

Dhyan Chandra Yadav, Saurabh Pal [14] The main aim of this research study was to detect thyroid disease using three different classification algorithms like Random Tree, J48 and Hoeffding. In this research dataset of 499 patient were used. Accuracy of individual algorithm is calculated. A new ensembled method were introduce by author and apply again on the same dataset which is used to calculate individual algorithm accuracy. From the experimental result it is observed that ensembled method provides better classification having 99.2% accuracy and 99.36% sensitivity.

Yijun Wu, Ke Rao et al. [15] The main aim of this study is to develop machine learning based predictive model to detect Central lymph node metasta sis (CLNM) which is occur frequently in patients with papillary thyroid cancer (PTC). In this study 1103 PTC patient data were reviewed. Six machines learn ing algorithm were applied in this study like AdaBoost (adaptive boosting), DT (decision tree), random forest classifier (RFC), ANN, extreme gradient boosting (XGBoost) and gradient boosting decision tree (GBDT). This study conclude GBT is best machine learning based model for the prediction of CLNM in PTC patient.

Jagdeesh saraf and Dr. Kalpana V. [16] A computer-based technique for segmenting and classifying the nodules as malignant or benign is proposed in this research paper. US thyroid image dataset were used in this research. Edge detection techniques were used to segment the image and ANN classifier were used to classify nodule. Thresholding and template matching shape based fea ture extraction method were used and textural features were extracted from US thyroid image.

Massoud Sokouti, Mohsen Sokouti and Babak Sokouti [17] In this study, the important roles of detecting and diagnosing cold nodules has been discussed since they are known as high risk of infected by the thyroid cancer. The image preprocessing and processing techniques including image enhancement, image seg mentation (thresholding) were applied. These techniques couldn't solely capable of determining the cold nodules in thyroid radioisotope images so, a hill climbing algorithm was applied in order to automatic determination of these regions as it was much simpler than other intelligent systems such as Fuzzy SVM, ANN, SVM and genetic algorithms.

M Kalaiyarasi et al. [18] This study proposed a different machine learning classification algorithm to classify benign and malignant tumor. In this study breast cancer patient dataset is used. Machine learning algorithm like K-Nearest Neighbor, Logistic Regression, Support Vector Machine are used in this study.

Polepogu Rajesh, Kunduru Umamaheswari [19] Objective of this paper is to provide a complete solution to diagnosis the suspicious thyroid region in the thyroid gland. An advanced method of completely automatic identification of thyroid nodule is proposed in this paper. Microscopic images undergo automated procedure to recognize a disorder in the thyroid. The proposed method shows good results and identifies the thyroid nodule present in the image with good level of accuracy.

Fu-sheng Ouyang [20] The purpose of this study was to compare the classi f ication performance of linear and nonlinear machine-learning algorithms for the evaluation of thyroid nodules using pathological reports as reference standard. It is observed that the performance of both linear and non-linear machine learning algorithm is almost equal.

Lay Khoon Lee et al. [21] This review paper gives the overview of segmen tation method. It also gives the idea about which segmentation method should applied on digital images like MRI, CT scan, 3D MRI, X-Ray, and US images. It discusses about advantages and disadvantages of various segmentation methods.

Ahmet Akba¸ s et al. [22] This research study uses machine learning tech nique having aim to improve performance in diagnosis of thyroid cancer. In this research BayesNet, NaiveBayes, SMO, Ibk and Random Forest classifiers were used for classification. Using WEKA data mining program author compares these classifiers by using different measures like accuracy, Kappa, Matthews Cor relation Coefficient (MCC), and ROC. Experimental result shows that Random Forest classifier performs better.

Yongfeng Wang et al. [23] This paper compared the classification perfor mance of radiomics and deep learning-based methods using ultrasound images with thyroid nodules. The comparison results show that the deep learning-based model shows better accuracy.

Chandan R et al. [24] have proposed a machine learning based model to detect thyroid disease. In this research accuracy of individual ML algorithm were calculated, KNN provide 93.84% accuracy, SVM gives 95.38% accuracy, ANN provides 75.38% accuracy, Decision tree provides 92.3% accuracy, with Logistic Regression 96.92% accuracy is obtained. From the experimental reason it is observed that Logistic Regression outperformed other classifiers.

Rebecca Smith-Bindman [25] This research study is about to find out the risk of thyroid cancer using thyroid ultrasound image. They used three US thyroid imaging attributes 1) Entirely solid composition, 2) microcalcifications, 3) size larger than 2cm on which they decide whether sample need to undergo for biopsy or not. This is helpful to reduce unnecessary biopsies.

K. Sumithra et al. [26] This research study is about to presents an overview of different image processing techniques. It also provides the idea about image processing tool which researcher can use for image

processing. Various image processing techniques has been reviewing such as image recognition, image restoration, image enhancement and image segmentation.

Nikita Singh et al. [27] In this research author has propose a segmentation method and also performed a comparative study of KNN, SVM and Bayesian classifier. Thyroid ultrasound images were used for this research. MATLAB version 7.7.0 software were used for image processing. Accuracy of individual classifier is calculated, and it is observed that SVM has 84.62% accuracy, KNN has 46.15% accuracy and Bayesian has 38.46% accuracy. From the result obtained from experiment it is observed that SVM shows best accuracy

5.5 Machine Learning Techniques (Convolutional Neural Network (CNN), MobileNetV2)

Author(s)	Methodology	Dataset Type	Highlights	Limitations/ Gaps Identified	Analysis/ Remarks
Wei et al. (2020)	Ensemble Deep Learning	Multicenter US Images	Improved generalization across datasets	Did not focus on severity-level classification	Effective for multi-source data, but lacks detailed grading analysis.
Zhou et al. (2020)	Deep Learning Radiomics	Static US Images	Enhanced feature extraction capability	Limited to binary benign/malignant classification	Radiomics improved interpretability but lacks severity differentiation.
Wang et al. (2020)	Radiomics vs Deep Learning	Static US Images	DL models outperformed radiomics	Did not address multi-class or severity prediction	Showed DL superiority, suggesting potential for extended classification tasks.
Dong et al. (2020)	MobileNet V2 Architecture	General Images	Lightweight, computationally efficient	Not optimized for medical image analysis	Demonstrated architecture efficiency, but medical domain adaptation required.
Koh et al. (2020)	Deep CNN	Thyroid US Images	Accurate benign vs malignant detection	No severity or subtype labeling	Reliable for binary diagnosis; lacks depth in diagnostic categorization.
Yu et al. (2020)	Deep Learning Detection	Thyroid US Images	Automated thyroid nodule detection	No grading or severity analysis	Automation achieved, but model limited to initial detection phase.
Bai et al. (2020)	Deep Learning Risk Stratification	Thyroid US Images	Introduced risk scoring for nodules	Did not map to explicit severity levels	Good clinical applicability, could integrate severity mapping.

Abdolali et al. (2020)	Deep CNN Detection	Thyroid US Images	Automated nodule detection	Absence of severity-level analysis	Useful for detection tasks; insufficient diagnostic granularity.
Sharifi et al. (2021)	Deep Learning on Thyroid US	Static US Images	Learned key discriminative features	Lacked evaluation for specific disease classes	Model interpretable but limited in class differentiation.
Rehman et al. (2021)	Deep Learning Fast Screening	Thyroid US Images	Rapid and efficient screening	Focused only on binary diagnosis	Efficient for real-time use, not for detailed severity assessment.
Gulzar (2023)	MobileNet V2 + Transfer Learning	Agricultural Images	Strong performance with small datasets	Non-medical domain; lacks clinical validation	Good cross-domain learning example; medical adaptation needed.
Riaz et al. (2023)	MobileNet V2 for Segmentation	CT Lung Images	Effective feature transfer using MobileNetV2	Used CT images, not US data	Promising for transfer learning; requires modality adaptation for US.

6. Hypothesis

- The null hypothesis states that machine learning does not significantly improve thyroid cancer detection accuracy compared to traditional methods.
- The alternative hypothesis proposes that machine learning significantly enhances accuracy, sensitivity, and specificity, offering a more reliable diagnostic tool. This will be tested using performance metrics and statistical significance tests.

7. Methodology

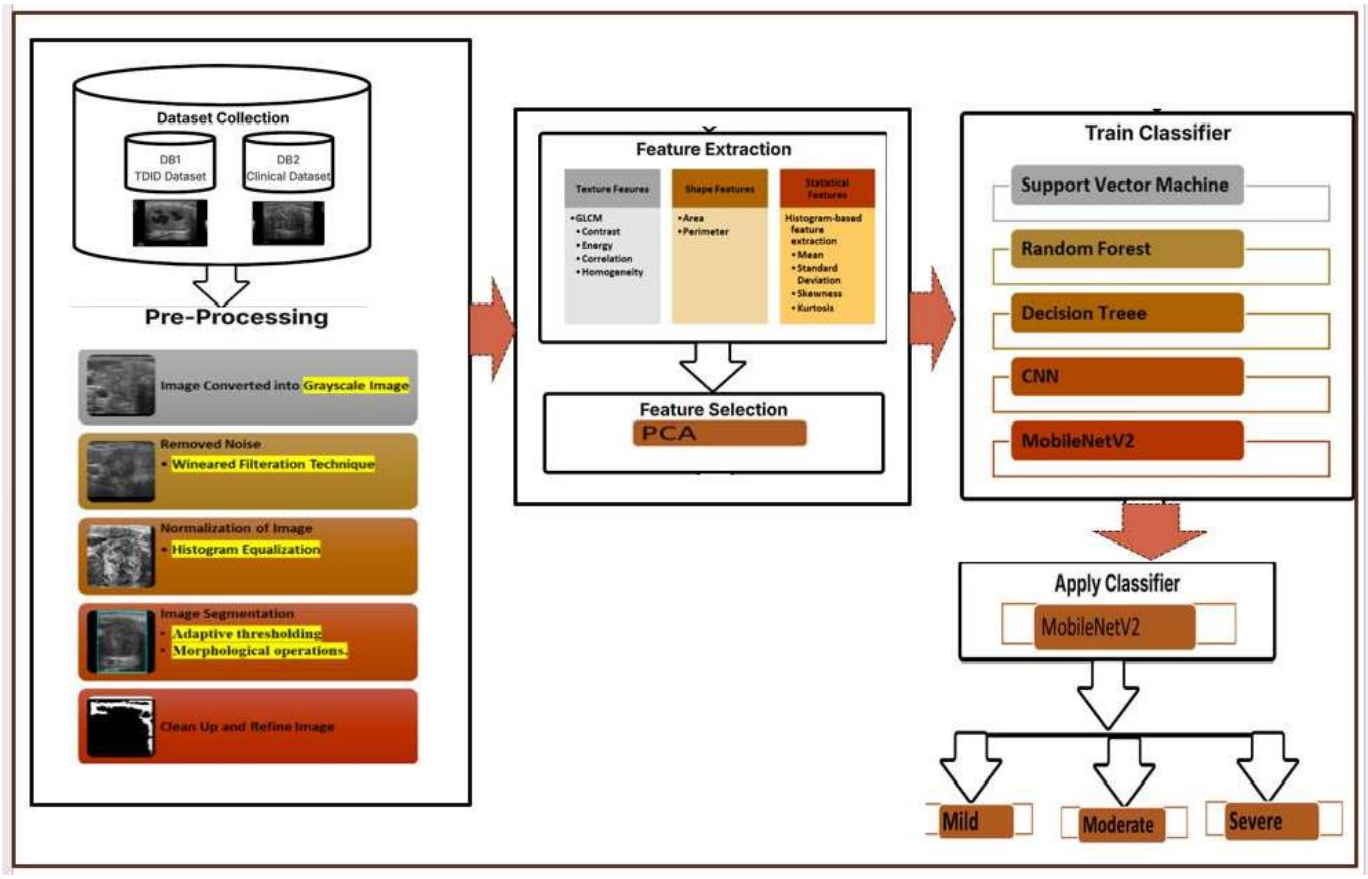


Figure 4 : Methodology

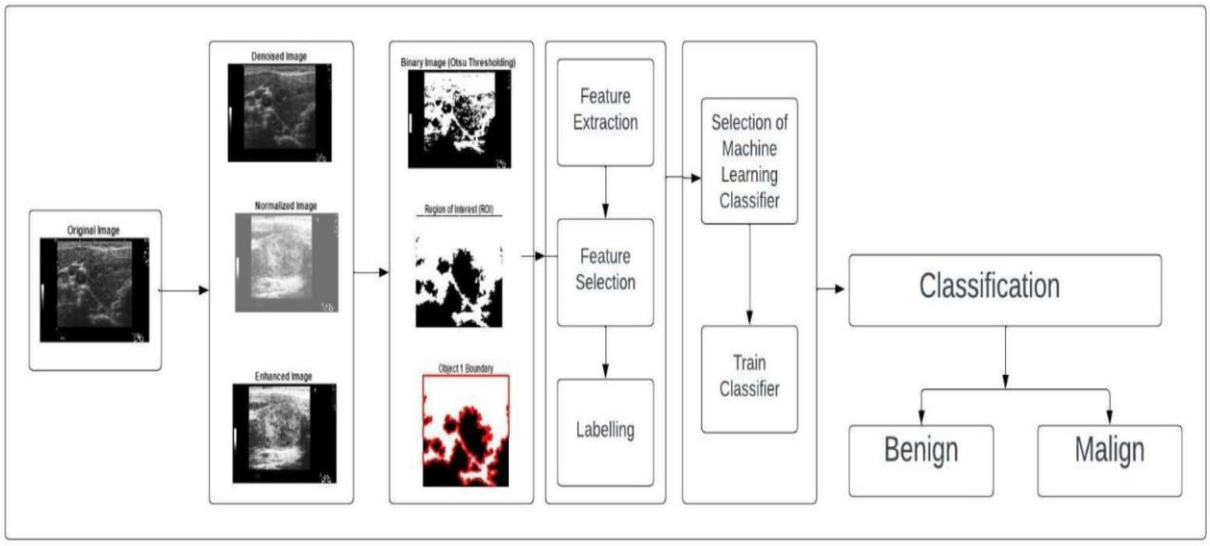


Figure 5 : Research Workflow

7.1 Dataset Collection:

Ultrasound imaging is one of the most commonly used diagnostic tools in the evaluation of thyroid nodules. It plays a crucial role in detecting abnormalities within the thyroid gland, as well as assessing the characteristics of nodules that may indicate malignancy. The process involves using high-frequency sound waves to create real-time images of the thyroid and surrounding structures. These images help physicians to distinguish between benign and suspicious nodules, although the distinction can sometimes be difficult without further intervention.

Database 1:

Dataset is collected from the University of Columbia database source-

This is a database of ultrasound images of thyroid nodules :

TDID (Thyroid Digital Image Database) is a freely accessible database of ultrasound images of thyroid nodules. Currently, this database has a group of B-mode ultrasound images (B-Mode is a two-dimensional ultrasound image display composed of bright dots representing the ultrasound echoes. The brightness of each dot is determined by the amplitude of the returned echo signal.), which include a complete annotation and diagnostic description of the suspicious images of thyroid lesions, made by expert radiologists.

Information of 389 patients has been collected.

More 99 patient information are also there which contain 33 Benign Images and 66 Malign Images.

Thyroid nodules are classified according to TIRADS based on their US features. Thyroid Imaging Reporting and Data Systems (TIRADS) accurately determine the risk of cancer in small thyroid nodules. In the TIRADS, the following four ultrasound features are scored 1 point each: irregular margins, hypoechogenicity (darkness), taller-than-wide shape, and microcalcifications; marked hypoechogenicity scored 2 points.

TIRADS 1 was defined as normal thyroid,
TIRADS 2 as the presence of benign features such as a cystic or spongiform nodule,
TIRADS 3 as the absence of suspicious features, and
TIRADS 4A as 1 point, TIRADS 4B as 2 points, TIRADS 4C as 3 to 4 points, and TIRADS 5 as 5 points, respectively.

Cancer rates were 0.9% in TIRADS 2, 2.9% in TIRADS 3, 12.3% in TIRADS 4A, 34.4% in TIRADS 4B, 66.6% in TIRADS 4C, and 86% in TIRADS 5

In our dataset TIRADS level in Benign cases,

22 images are of TIRADS 2 cases.

11 images are of TIRADS 3 cases.

TIRADS level in Malign cases,

20 images are of TIRADS 4a cases

16 images are of TIRADS 4b cases

19 images are of TIRADS 4c cases

11 images are of TIRADS 5 cases

This image information is in .xml file format.

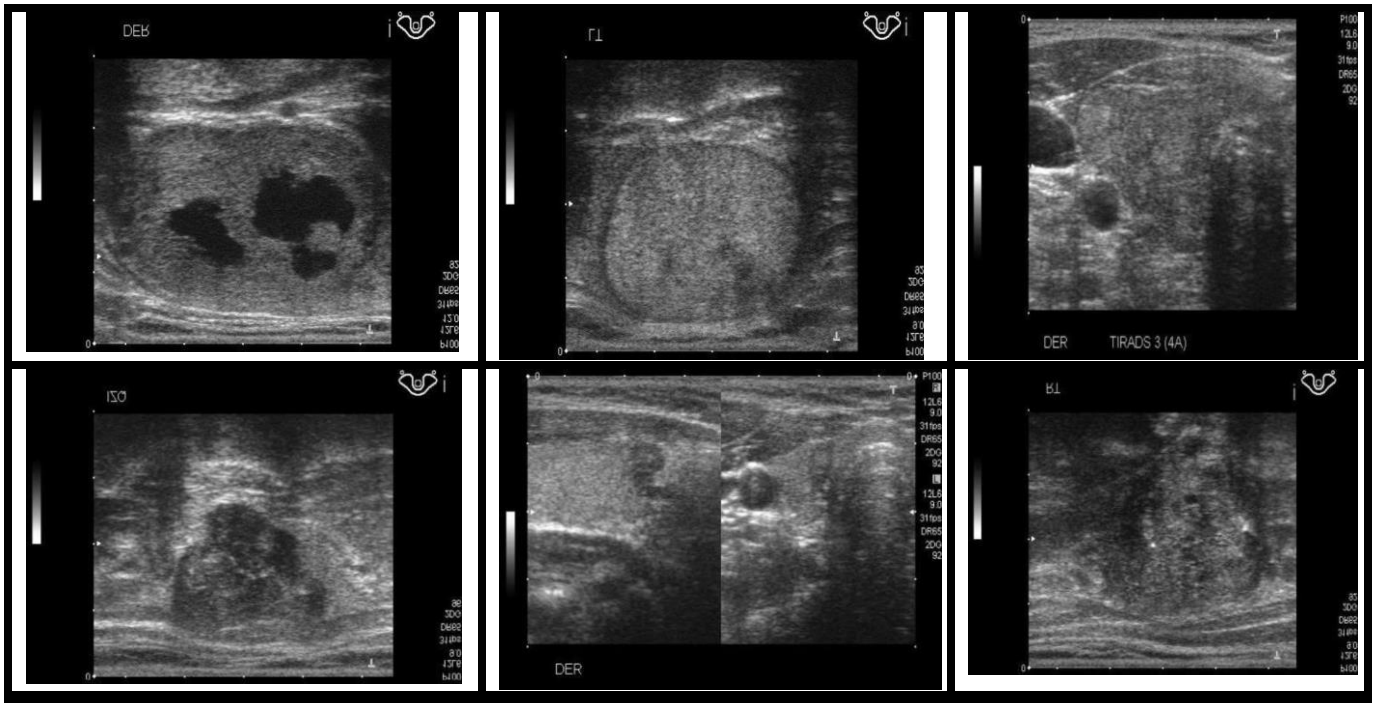


Figure 6: Sample Images (TDID Dataset)

Database 2:

The dataset used in this research is a cleaned dataset of 82 thyroid ultrasound images labeled by their severity as Mild, Moderate, or Severe. These images have been collected from an organized internal data repository, preserving uniform imaging protocols as well as labeling conventions. The severity labels were associated using corresponding descriptions found in the Severity_Classification_DDTI.xlsx dataset. This dataset features two main columns: "Image_ID and Severity_Label", enabling automatic linking of every ultrasound image to their corresponding severity label via programmatic preprocessing operations.

Each image was saved in JPEG format, with filenames being associated with the "Image_ID" field in the spreadsheet in an orderly manner. In order to achieve efficient MATLAB loading and labeling, an ImageDatastore was used, whereby the images, along with their respective severity labels, were consistently arranged.

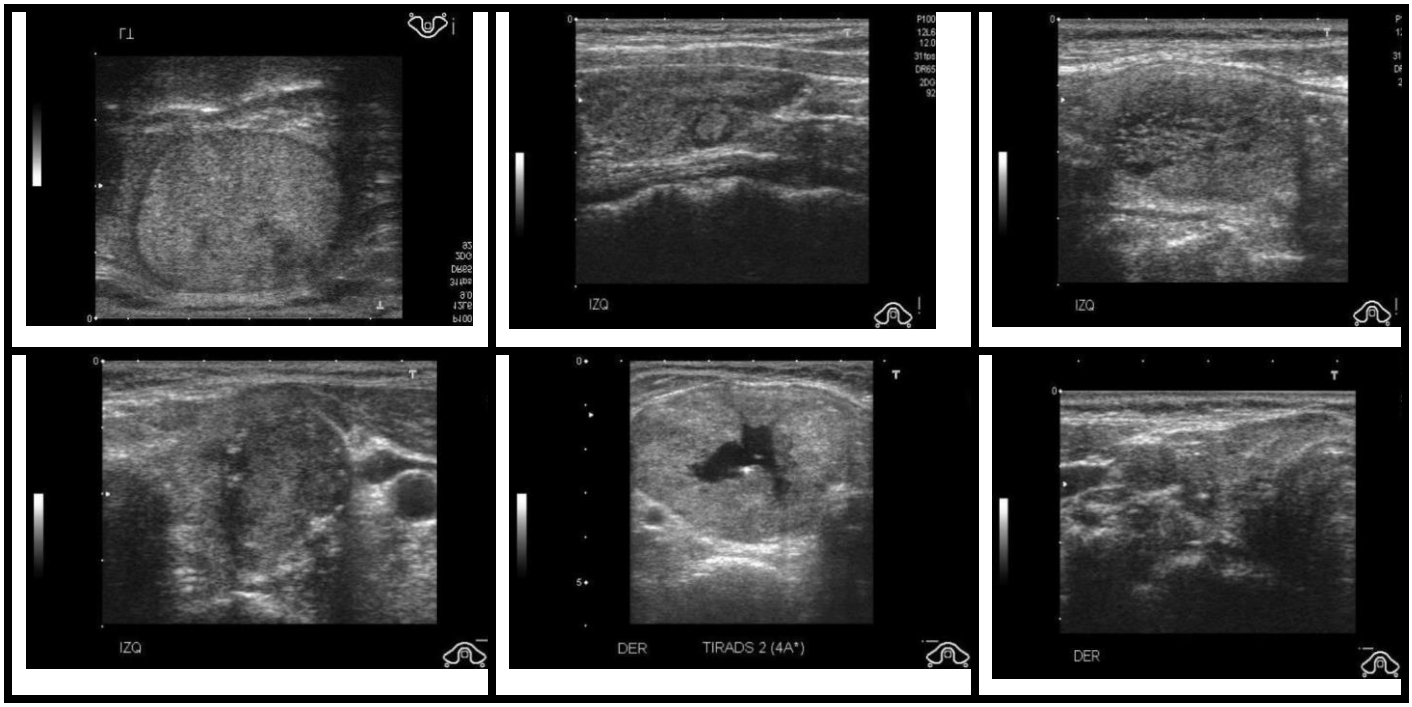


Figure 7:Sample Images (Clinical Dataset)

7.2 Image Preprocessing:

Figure 8 :Preprocessing Steps

Step 1: Load Ultrasound Image

Step 2: Converted into Gray scale image.

Grayscale images contain only shades of gray and do not have color information. Instead, they represent the intensity of each pixel, typically ranging from black (lowest intensity) to white (highest intensity). It simplifies data processing and analysis, making it more computationally efficient. In some feature extraction algorithms (e.g., edge detection, texture analysis), working with grayscale images simplify the calculations and enhance the detection of important features.

Step 3: Removed noise from Gray scaled image.

Using median filter. Employing a Median filter enhances image quality and boosts the precision of diagnosis and analysis, thereby improving overall effectiveness. The median filter is a type of spatial domain filtering technique commonly used for noise reduction in images. It works by replacing each pixel value with the median value of its neighboring pixels within a defined window. It preserves edge

information, reduces paper and salt noise from image, it also retains image details. The median filter provides noise reduction while preserving the essential image details.

Step 4: Normalization of image using histogram equalization.

Histogram equalization is a widely used image enhancement technique that involves redistributing the intensity values of pixels across a wider range. This process aims to achieve a more balanced and visually appealing representation of the image. Specific purposes and benefits of normalization using histogram equalization are contrast enhancement, Better Visualization, Improved Feature Detection.

Step 5: Image segmentation using Otsu's method for thresholding.

Otsu's method is a widely used technique for image segmentation. It separates an image into two classes, foreground, and background, based on the grayscale intensity values of its pixels. Furthermore, Otsu's method uses the grayscale histogram of an image to detect an optimal threshold value that separates two regions with maximum inter-class variance. It was proposed by Nobuyuki Otsu in 1979 and has since become widely used due to its simplicity and effectiveness. The main goal of Otsu's Method is to find the threshold that maximizes the between-class variance, while minimizing the within-class variance. The method assumes that the image contains two main classes with distinct intensity levels, and the threshold is sought to separate these classes effectively.

Step 6: To clean up and refine the binary image after thresholding or segmentation.

In image processing, after thresholding or segmentation, it's often necessary to clean up and refine the binary image to remove noise, fill gaps, and make the objects of interest more well-defined.

7.3 Feature Extraction:

Figure 9 :Feature Extraction Methods

The image feature extraction plays an important role in the detection of thyroid cancer. After completing image segmentation, features are extracted from the image to detect thyroid cancer. Feature extraction is an essential stage that rep resents the final results to predict cancer and non-cancer of an image. Following methods are used to extract feature from pre-processed image.

1. Gray-Level Co-occurrence Matrix (GLCM)

2. Intensity based

3. Histogram based

7.3.1. Gray Level Cooccurrence Matrix (GLCM):

In image classification, the Gray Level Cooccurrence Matrix (GLCM) is a frequently used feature ex traction approach. GLCM represents the second-order statistical information of gray levels between neighboring pixels in an image. It calculates the cooccurrence of neighboring gray levels and their counts in an image using a square matrix based on the ROI dimension of the number of gray levels in the US-thyroid image. Energy: The energy feature in an image measures the degree of pixel pair repetitions and texture disorder, with highly correlated pixels resulting in a large energy value.

'Energy' Is defined using an equation.

$$\text{Energy} = \sum_{i,j} p(i,j)^2$$

Entropy: The entropy feature assesses the uniformity of pixels within an image and employs randomness to describe the image's texture. Entropy and energy are inversely correlated, meaning that images with a greater number of gray levels tend to have higher entropy.

'Entropy' is defined using an equation.

$$\text{Ent} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j) \log(p(i,j))$$

Homogeneity: Homogeneity is often described as the inverse of different movement. It quantifies the consistency of an image by yielding higher values when there is minimal variation in gray tones between paired objects. In contrast to energy, which remains constant, homogeneity diminishes as contrast in the image increases.

'Homogeneity' is defined using an equation.

$$Homog = \sum_i \sum_j \frac{1}{1+|i-j|^2} P(i,j)$$

Contrast: Contrast quantifies the spatial frequency of an image, specifically the differences in pixel intensity between a pixel and its neighbors across the entire image. It serves to assess the extent of local variations in the image, highlighting distinctions in color and brightness between objects within the same field of view.

'Contrast' is defined using an equation.

$$Cont. = \sum_i \sum_j |i-j|^2 p(i,j)$$

Correlation: The correlation feature of an image measures the pixel's correlation with its neighbor over the entire image. It ranges from -1 to 1, with negative correlations being perfectly positive and constant images having infinity values. 'Correlation' Is defined using an equation.

$$\text{Correl} = \sum_i \sum_j \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i \sigma_j}$$

Angular Second Moment (ASM): The Angular Second Moment (ASM), also known as Uniformity or Energy, is a measure of image homogeneity, calculated by adding squares of entries in the GLCM. High ASM indicates good homogeneity or similar pixels. 'ASM' is defined using an equation.

$$\text{ASM} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} P_{ij}^2$$

7.3.2 Shape based feature extraction:

Shape-based feature extraction is a technique in image processing and computer vision where meaningful features related to the geometry or structure of objects in an image are extracted. These features are often used in tasks like object recognition, image classification, medical imaging, biometrics, and more.

Common Shape-Based Features

i. Contour-based Features:

Perimeter: Length of the object boundary.
Eccentricity: Shape elongation (ratio of major to minor axes).
Convex Hull: The smallest convex shape that can enclose the object.
Solidity: Ratio of object area to convex hull area.

ii. Region-based Features:

Area: Total number of pixels inside the shape.
Centroid: Geometric center of the object.
Extent: Ratio of object area to bounding box area.
Equivalent Diameter: Diameter of a circle with the same area as the object.

7.3.3 Histogram-based feature extraction:

In image processing, intensity-based feature extraction methods analyze the pixel intensities of an image without considering the arrangement of pixels. Some commonly used intensity-based features include:
Mean Intensity: The average intensity value of all pixels in the image.
Mean Intensity = (Sum of all intensity values) / (Total number of intensity values)
Median Intensity: The middle value of the sorted intensity values.
Standard Deviation: A measure of the spread of intensity values from the mean intensity.
Histogram: A graphical representation of the intensity distribution in the image.
Percentiles: The value below which a given percentage of intensity values fall.
Skewness and Kurtosis: Measures of the asymmetry and peakedness of the intensity distribution, respectively.

7.4 Selection of Features:

After extracting features using the Gray Level Co-occurrence Matrix (GLCM), applying PCA is important because it reduces the high dimensionality and redundancy of the feature set. GLCM features often contain correlated information, and PCA transforms them into a smaller set of uncorrelated components that capture the most important variance in the data. This simplification improves computational efficiency, reduces noise and overfitting, and often enhances the performance of machine learning models.

7.4.1 Principal Component Analysis (PCA):

It is a technique used in statistics and machine learning for dimensionality reduction. The primary goal of PCA is to transform high-dimensional data into a lower-dimensional representation while retaining as much of the original information as possible. This is achieved by identifying the principal components, which are linear combinations of the original features that capture the maximum variance in the data.

Working of PCA:

Standardization: Before applying PCA, it's common to standardize the data by subtracting the mean and dividing by the standard deviation. This ensures that all features have a similar scale. Covariance Matrix: PCA computes the covariance matrix of the standardized data. The covariance matrix represents the relationships between different features, indicating how they vary together. Eigenvalue Decomposition: The next step involves finding the eigenvalues and eigenvectors of the covariance matrix. Eigenvalues represent the variance explained by each principal component, and eigenvectors represent the direction of maximum variance. Principal Components: The eigenvectors corresponding to the highest eigenvalues are the principal components. These components are used to form a new basis for the data. Dimensionality Reduction: The final step is to project the original data onto the new basisformed by the principal components. By selecting a subset of the principal components, one can reduce the dimensionality of the data.

7.5 Apply Classification Algorithm:

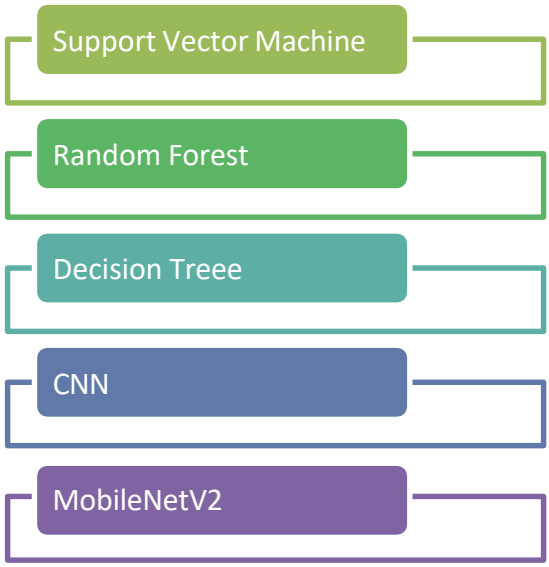


Figure 10: Classification Algorithm

7.5.1. Support Vector Machine (SVM):

Support Vector Machine is a supervised machine learning algorithm used for classification tasks. It finds the best boundary (called a hyperplane) that separates different classes (e.g., cancerous vs. non-cancerous) with the largest margin. The points closest to this boundary are called support vectors, which help define the position of the hyperplane.

How SVM works to detect thyroid cancer:

- Since SVM works with numeric feature vectors, raw ultrasound images are first transformed into meaningful features (e.g., texture, edges, intensity patterns) using image processing or a pre-trained CNN.
- These features are then fed into the SVM.
- SVM learns the boundary that best separates cancerous from non-cancerous thyroid images based on these features.
- When a new ultrasound image is presented, its features are extracted and classified using this boundary to predict if it shows signs of thyroid cancer.

7.5.2. Random Forest:

Random Forest is an ensemble learning method that combines many decision trees. Each tree is trained on a random subset of the data and features, and they vote to decide the final class. This reduces overfitting compared to a single decision tree and improves accuracy.

How Random Forest works to detect thyroid cancer:

- Similar to SVM, Random Forest requires image features as input. These can be handcrafted or extracted using a CNN.
- Multiple decision trees are built, each learning different patterns that distinguish cancerous vs. healthy tissue in ultrasound images.
- The ensemble aggregates the predictions from all trees to classify images more robustly.
- For a new image, its features are extracted, run through all trees, and the majority vote predicts cancer presence or absence.

7.5.3. Decision Tree:

A Decision Tree is a simple flowchart-like structure where each internal node tests a feature, branches represent test outcomes, and leaves represent final decisions (classes). It recursively splits the data based on features to separate classes.

How Decision Tree works to detect thyroid cancer:

- Features from ultrasound images are extracted beforehand.
- The tree learns rules like "if texture measure > threshold, then cancerous; else non-cancerous" by splitting the feature space.
- For a new image, its features are checked against the learned rules to classify it.
- It's easy to interpret but can overfit the training data if not pruned properly.

7.5.4. Convolutional Neural Network. (CNN):

A Convolutional Neural Network (CNN) model designed to automatically extract and learn features from medical images. It is particularly effective for image-based disease detection and classification.
In the context of thyroid cancer detection, CNN analyzes ultrasound or histopathological images to distinguish between benign and malignant nodules.
The process involves several key layers:
Convolution Layer:This layer applies filters to the input image to detect important features such as edges, textures, and patterns related to thyroid tissue structure.
Pooling Layer:It reduces the dimensionality of the feature maps, keeping the most significant features while minimizing computational complexity.
Activation Function (ReLU):The Rectified Linear Unit introduces non-linearity, enabling the model to learn complex patterns in the image data.
Fully Connected Layer:This layer integrates the extracted features and classifies the image into categories (e.g., benign or malignant).
Output Layer:Produces the final prediction, indicating the probability of the image belonging to each class.
By training on a large set of labeled thyroid images, the CNN learns to recognize subtle differences between healthy and cancerous tissues, providing high diagnostic accuracy and consistency compared to manual evaluation.

7.5.5. MobileNetV2:

MobileNetV2 is a convolutional neural network architecture designed for efficient, lightweight image analysis. CNNs process images through layers that automatically learn hierarchical features—edges, shapes, textures—directly from raw pixels.

How MobileNetV2 works to detect thyroid cancer:

- The raw ultrasound images are input directly into the network.
- The network automatically learns relevant features during training that distinguish cancerous tissues from healthy ones.
- It fine-tunes convolutional layers to capture subtle differences in image texture and structure indicative of thyroid cancer.
- Once trained, MobileNetV2 can classify new ultrasound images with high accuracy without manual feature extraction.

7.6 Train Classifier:

In this research work, a machine learning classifier is trained to automatically categorize thyroid nodules into three classes: mild, moderate, and severe, based on extracted features from medical data such as ultrasound images and patient records. The training process involves feeding the algorithm with a labeled dataset where each instance contains measurable features (e.g., shape irregularity, margin sharpness, texture scores, or other clinical parameters) along with its known classification. The classifier learns patterns and correlations within the data that distinguish one class from another. By using supervised learning algorithms such as Random Forest, Support Vector Machine (SVM), or Logistic Regression, the model is able to learn the underlying structure of the data. Once trained, the classifier can be used to accurately predict the severity level of new, unseen thyroid nodules, thereby assisting medical

professionals in risk assessment, diagnosis, and treatment planning. The ultimate goal is to provide a reliable, fast, and non-invasive tool for early detection and classification of thyroid cancer severity.

7.7 Classify Nodule into Malign or non-malign:

To enable accurate detection of thyroid cancer, in this research a machine learning classifier is trained to categorize thyroid nodules as either benign (non-cancerous) or malignant (cancerous) based on extracted features from medical data, primarily ultrasound images and clinical parameters. The training process involves providing the algorithm with a labeled dataset where each nodule is described using key features such as shape, size, margin irregularity, texture, and echogenicity, along with its confirmed diagnosis. The classifier learns to identify patterns and relationships in the data that distinguish benign nodules from malignant ones. Supervised learning algorithms such as Random Forest, Support Vector Machine (SVM), or Logistic Regression are used to build the predictive model. Once trained, the classifier can be applied to new, unseen data to assist radiologists and clinicians in making more accurate, consistent, and timely diagnoses. This approach aims to support early detection of thyroid cancer, reduce unnecessary biopsies, and improve patient outcomes through data-driven decision-making.

8. Experiment and Results:

The experiments involved preprocessing the ultrasound images, extracting relevant features, training the models, and evaluating their performance using standard metrics such as accuracy, precision, recall, and F1-score. Statistical analysis was also performed to validate the significance of the results.

To evaluate the effectiveness of machine learning techniques in detecting thyroid cancer from ultrasound images, a series of experiments were conducted.

This research work involved training and evaluating three different models for the classification of thyroid nodules as benign or malignant:

- Support Vector Machine (SVM)
- Convolutional Neural Network (CNN)
- MobileNetV2 using transfer learning

The goal was to compare the performance of traditional feature-based and deep learning-based approaches in detecting thyroid cancer from ultrasound images.

8.1 Dataset:

This research utilizes two distinct datasets for training, testing, and validating machine learning models for thyroid nodule classification:

TDID (Thyroid Digital Image Database)

A public dataset from the University of Columbia containing B-mode ultrasound images of thyroid nodules. It includes 488 patient records, with detailed annotations and TIRADS-based classification (ranging from TIRADS 2 to 5), allowing for binary classification of nodules as benign or malignant.

Internal Clinical Dataset

A set of 82 ultrasound images collected from a clinical repository, labeled as Mild, Moderate, or Severe based on expert diagnosis. Images were organized using MATLAB's ImageDatastore, with labels linked via a structured Excel file.

Both datasets were used to train and validate ML models, enabling comparison across binary and multi-class classification tasks.

8.2 Preprocessing

Before feature extraction and model training, exhaustive preprocessing was done to improve image quality in the ultrasound images as well as to maintain standardization in the dataset. The preprocessing chain included the following sequentially executed processes:

8.2.1 Grayscale Conversion:

All ultrasound images input were confirmed to be modality consistent. While source images were mainly grayscale, every image was intentionally converted to grayscale by means of MATLAB's "rgb2gray" system if needed. This process created image uniformity in channels and reduced unnecessary information for compliance with standard requirements by subsequent enhancement methods.

8.2.2 Median Filtering:

Ultrasound images naturally have speckle noise, which masks important anatomical features. This was reduced by applying the function "medfilt2" to perform 3×3 median filtering. Median filtering efficiently retained edge details by eliminating random noise, resulting in enhanced boundary clarity for nodules as well as internal texture features.

8.2.3 Contrast Enhancement:

In order to further increase contrast between the nodule structures and other structures in the sonograms, Contrast Limited Adaptive Histogram Equalization (CLAHE) was utilized through MATLAB's "adapthisteq" function. CLAHE adapts contrast in regions dynamically in an image, thus enhancing subtle intensity variations that are very important for proper feature extractions. This greatly improved local contrast, especially in areas with limited contrast common in thyroid sonograms.

8.2.4 Region of Interest (ROI) Segmentation:

After reduction of noises and contrast enhancement, Region of Interest (ROI) holding the thyroid nodule was segmented. In adaptive thresholding, we used the "imbinarize" function with the 'adaptive' approach, which estimates thresholds from local mean intensity. Morphological operations such as opening (imopen) and closing (imclose) with disk structuring elements were applied after thresholding to improve the segmented binary mask. The largest connected component was selected by utilizing "bwareafilt", assuming that it was the nodule area. This process efficiently separated the nodule from background tissues and artifacts. Optional computation of bounding boxes was done on "regionprops" for graphical verification of segmented ROIs.

8.2.5 Augmentation for Training:

Data augmentation methods were incorporated in the training process to compensate for the small dataset size to achieve better model generalization. These data augmentations included: Random Rotation (±10 degrees), Horizontal/Vertical Translation (translations by 5 pixels), and Random Scaling (random zoom by ±10%). Data augmentation was done with MATLAB's "imageDataAugmenter" class to keep the augmented images clinically representative while adding variability to support strong training.

Benign Image-

In this experiment, we have applied preprocessing steps to improve image quality.

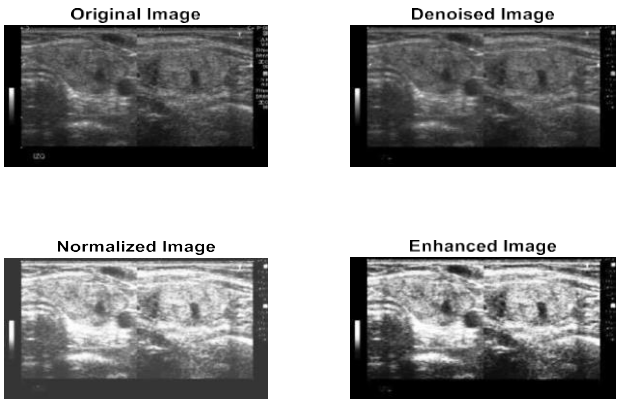


Figure 11.(A): Benign Pre-processed image

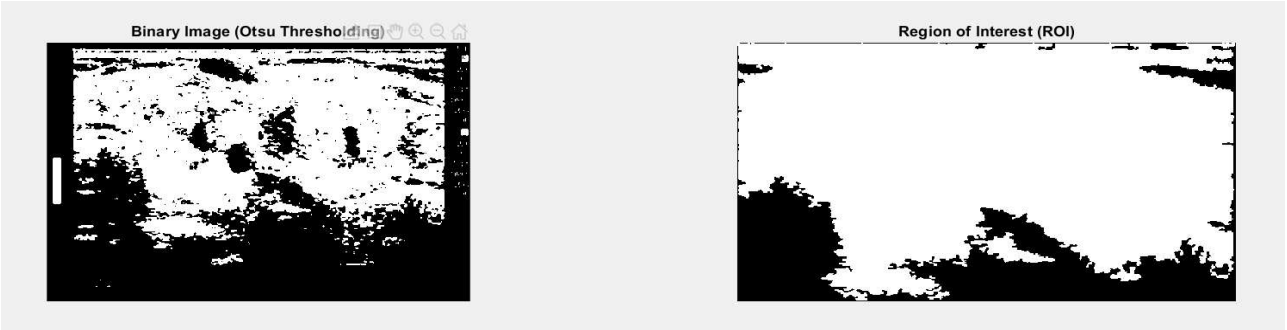


Figure 11(B): Benign Segmented image

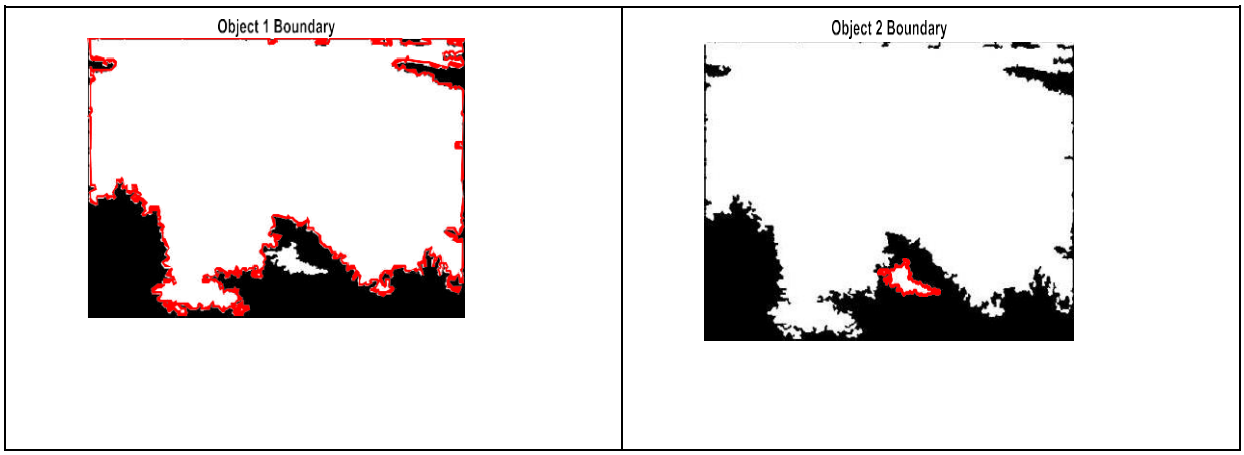


Figure 11(C): Benign Object Boundary image

Malign Image:

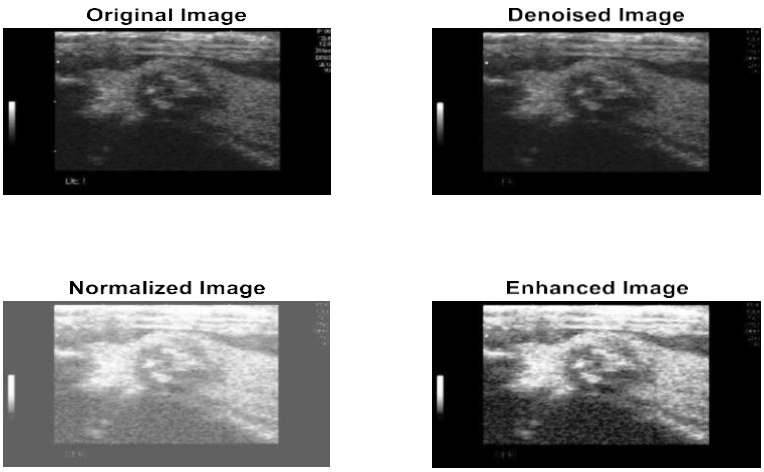


Figure 12(A) Malign Pre-processed Image

Figure 12(B): Malign Segmented Image

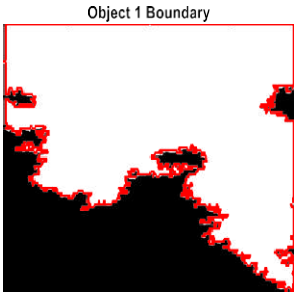


Figure 12(C): Malign Object boundary Image

8.3 Feature Extraction

8.3.1. Support Vector Machine:

For Support Vector Machine, After preprocessing, three types of feature extraction techniques were applied:

- **Shape-based features** were computed to describe the geometric structure of the nodule, including parameters such as **area**, **perimeter**, **compactness**, **eccentricity**, and **solidity**. These features help in understanding the contour and structure of the nodules, which can differ significantly between benign and malignant cases.
- **GLCM (Gray-Level Co-occurrence Matrix) features** were extracted to capture the **texture patterns** within the nodule. GLCM-based features such as **contrast**, **correlation**, **homogeneity**, **energy**, and **entropy** provide insights into the spatial arrangement of pixel intensities and are effective in distinguishing nodules based on their internal texture.
- **Intensity-based features** represent statistical measures of pixel intensity values within the nodule region, including **mean**, **and standard deviation**. These features help to evaluate the brightness and intensity distribution, which can indicate tissue abnormality.

Table 1: GLCM Features

Features	Image1 Normal	Image2 Normal	Image3 Normal	Image4 Abnormal	Image5 Abnormal	Image6 Abnormal	Image7 Abnormal	Image8 Abnormal	Image9 Abnormal	Image10 Abnormal
contrast_1	0.018	0.011	0.014	0.014	0.012	0.009	0.012	0.013	0.013	0.010
contrast_2	0.037	0.024	0.028	0.027	0.022	0.022	0.024	0.028	0.023	0.023
contrast_3	0.030	0.020	0.022	0.023	0.017	0.019	0.019	0.022	0.017	0.017
contrast_4	0.036	0.024	0.026	0.028	0.022	0.022	0.023	0.027	0.023	0.022
energy_1	0.545	0.555	0.588	0.603	0.538	0.605	0.554	0.591	0.583	0.659
energy_2	0.529	0.544	0.576	0.592	0.529	0.596	0.545	0.579	0.576	0.651
energy_3	0.534	0.547	0.580	0.595	0.533	0.597	0.548	0.582	0.581	0.654
energy_4	0.529	0.544	0.578	0.592	0.529	0.596	0.546	0.579	0.576	0.652
correlation_1	0.957	0.975	0.965	0.965	0.974	0.977	0.971	0.967	0.967	0.968
correlation_2	0.914	0.944	0.928	0.929	0.950	0.942	0.945	0.930	0.943	0.930
correlation_3	0.929	0.954	0.944	0.941	0.963	0.952	0.957	0.944	0.958	0.948
correlation_4	0.915	0.945	0.934	0.927	0.951	0.943	0.948	0.931	0.942	0.934
homogeneity_1	0.990	0.995	0.993	0.993	0.994	0.995	0.994	0.993	0.993	0.995
homogeneity_2	0.981	0.988	0.986	0.986	0.989	0.989	0.988	0.986	0.988	0.989
homogeneity_3	0.984	0.990	0.989	0.989	0.992	0.991	0.991	0.989	0.992	0.991
homogeneity_4	0.981	0.988	0.987	0.986	0.989	0.989	0.989	0.986	0.988	0.989

Table 2: Histogram Based Features

Features	Image1 Normal	Image2 Normal	Image3 Normal	Image4 Abnormal	Image5 Abnormal	Image6 Abnormal	Image7 Abnormal	Image8 Abnormal	Image9 Abnormal	Image10 Abnormal
Mean	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Variance	0.063	0.065	0.100	0.115	0.048	0.112	0.065	0.102	0.095	0.167
Kurtosis	1	1	1	1	1	1	1	1	1	1
Percentile25	0.323	0.320	0.277	0.261	0.345	0.263	0.320	0.274	0.282	0.211
Percentile50	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Percentile75	0.677	0.680	0.723	0.739	0.655	0.737	0.680	0.726	0.718	0.789
Entropy	1	1	1	1	1	1	1	1	1	1

Table 3: Intensity Based Features

Features	Img1 Normal	Img2 Normal	Img3 Normal	Img4 Abnormal	Img5 Abnormal	Img6 Abnormal	Img7 Abnormal	Img8 Abnormal	Img9 Abnormal	Img10 Abnormal
mean_ intensity	0.677	0.680	0.723	0.739	0.655	0.737	0.680	0.726	0.718	0.789
std_ Deviation	0.468	0.467	0.447	0.439	0.475	0.440	0.466	0.446	0.450	0.408

8.3.2 Convolutional Neural Network (CNN)

Effective feature extraction is essential for accurate classification of thyroid nodules from ultrasound images. In this research, the Convolutional Neural Network (CNN) model leverages a combination of advanced image processing techniques—Gray-Level Co-occurrence Matrix (GLCM), Discrete Wavelet Transform (DWT), and Histogram of Oriented Gradients (HOG)—to extract rich textural and spatial features.

GLCM captures second-order texture statistics by analyzing pixel intensity relationships. Features like contrast, correlation, energy, and homogeneity help distinguish benign from malignant nodules.

DWT decomposes the image into multiple frequency sub-bands, enabling the detection of fine structural variations and subtle edge details within the nodules.

HOG captures local shape and edge orientation information by analyzing the distribution of gradient directions, which enhances the model's ability to detect nodular boundaries and shape irregularities.

The extracted features serve as input for classification models, ensuring that both texture and spatial variations in ultrasound images contribute to the learning process. This combination of techniques enhances the robustness of the proposed system in distinguishing thyroid abnormalities.

8.3.3. MobileNetV2

After preprocessing, we used a pretrained MobileNetV2 architecture to perform deep feature extraction. The enhanced ultrasounds were mapped into a discriminative representation in a reduced dimension using feature extraction.

8.3.3.1 Selection of Feature Extractor

MobileNetV2, being a light-weight convolutional neural network pre-trained on ImageNet dataset, was chosen as the extractor backbone. Its linear-bottleneck inverted structure facilitates efficient learning of representations, making it very suitable for small datasets as well as for medical image tasks where efficiency in computations and generality are crucial. MATLAB's "mobilenetv2" function loaded the network, and all original fully connected layers, the softmax layer, as well as the classification layer, were sequentially stripped off using MATLAB's "layerGraph" functions. The stripping left behind fully convolutional layers till the global average pooling layer that captures spatial features in the image.

8.3.3.2 Feature Map Extraction

The Global Average Pooling 2D layer (global_average_pooling2d_1) produced output that was used as the feature vector for every individual image. This layer pools spatial information over width and height dimensions to transform the feature map to one 1280-dimensional feature vector per image. This succinct but informative representation enabled efficient learning in later stages of classification.

The feature activations were retrieved through MATLAB's "activations" function with the parameters set to have the features as row vectors. The extracted features set X was organized

$$X = \{x_1, x_2, \dots, x_n\} \quad where \quad x_i \in R^{1280}$$

where n is the number of images.

8.4 Model Training

1. Three popular supervised learning algorithms were employed in this research:

1. **Support Vector Machine (SVM)**: This classifier is effective in high-dimensional spaces and works by finding the optimal hyperplane that best separates the classes. It is particularly useful for binary classification tasks like this one, and it performs well even with non-linear data when combined with kernel functions.
2. **Random Forest**: An ensemble learning method that builds multiple decision trees during training and outputs the class that is the mode of the classes predicted by individual trees. Random Forest is robust to overfitting, handles large datasets well, and can capture complex interactions among features.
3. **Decision Tree Classifier**: A simple yet powerful model that splits the data based on feature thresholds to create a tree-like structure. It is easy to interpret and visualize, making it useful for understanding how the model makes decisions.

   Each classifier was trained on the same dataset to evaluate and compare their performance in terms of accuracy, precision, recall, and F1-score. This multi-model approach helps determine the most effective algorithm for reliable and automated thyroid cancer detection based on medical image analysis.

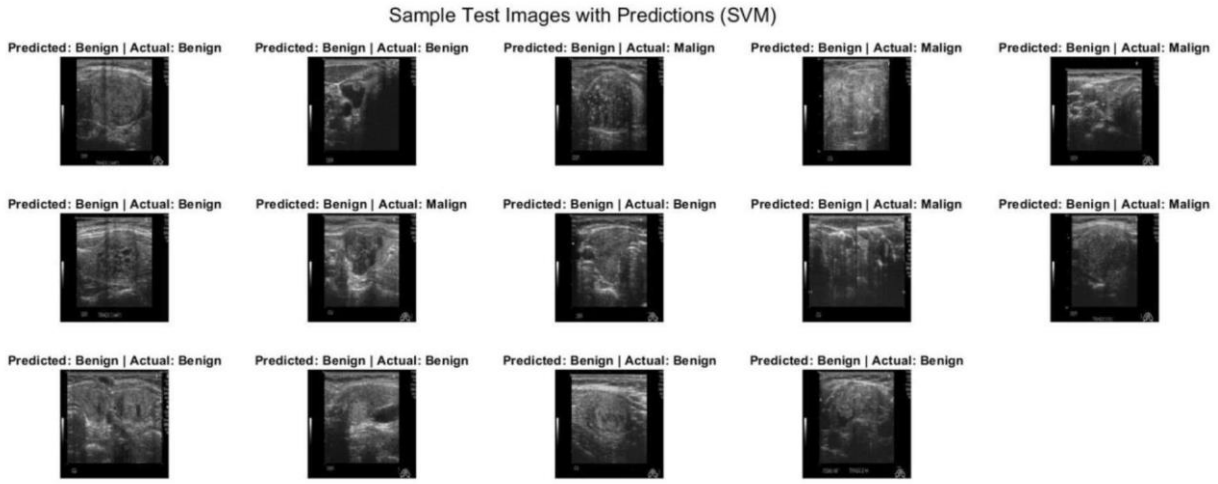

Sample Test Images with Predictions (SVM)
Predicted: Benign | Actual: Benign
Predicted: Benign | Actual: Benign
Predicted: Benign | Actual: Malign
Predicted: Benign | Actual: Malign
Predicted: Benign | Actual: Malign
Predicted: Benign | Actual: Benign
Predicted: Benign | Actual: Malign
Predicted: Benign | Actual: Benign
Predicted: Benign | Actual: Malign
Predicted: Benign | Actual: Malign
Predicted: Benign | Actual: Benign
Predicted: Benign | Actual: Benign
Predicted: Benign | Actual: Benign
Predicted: Benign | Actual: Benign

Figure 13 A: Sample Test Images with Predictions (SVM)

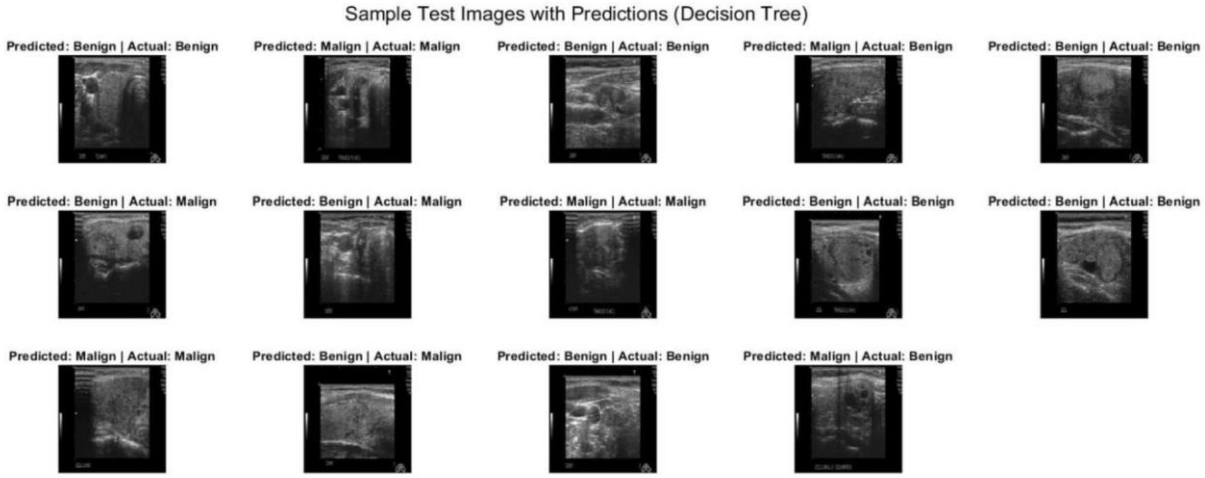

Sample Test Images with Predictions (Decision Tree)
Predicted: Benign | Actual: Benign
Predicted: Malign | Actual: Malign
Predicted: Benign | Actual: Benign
Predicted: Malign | Actual: Benign
Predicted: Benign | Actual: Benign
Predicted: Benign | Actual: Malign
Predicted: Benign | Actual: Malign
Predicted: Malign | Actual: Malign
Predicted: Benign | Actual: Benign
Predicted: Benign | Actual: Benign
Predicted: Malign | Actual: Malign
Predicted: Benign | Actual: Malign
Predicted: Benign | Actual: Benign
Predicted: Malign | Actual: Benign

Figure 13 B : Sample Test Images with Prediction (Decision Tree)

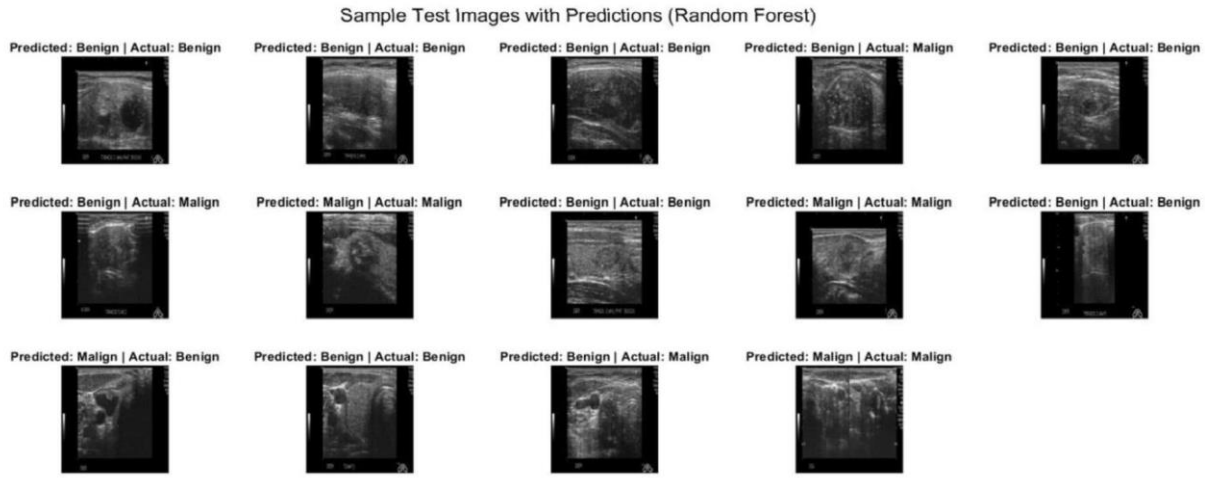

Sample Test Images with Predictions (Random Forest)
Predicted: Benign | Actual: Benign
Predicted: Benign | Actual: Benign
Predicted: Benign | Actual: Benign
Predicted: Benign | Actual: Malign
Predicted: Benign | Actual: Benign
Predicted: Benign | Actual: Malign
Predicted: Malign | Actual: Malign
Predicted: Benign | Actual: Benign
Predicted: Malign | Actual: Malign
Predicted: Benign | Actual: Benign
Predicted: Malign | Actual: Benign
Predicted: Benign | Actual: Benign
Predicted: Benign | Actual: Malign
Predicted: Malign | Actual: Malign

Figure 13 C: Sample Test Images with Prediction (Random Forest)

Table 4: Performance Evaluation Parameters

Model	Precision	Recall	F1-score	Test Accuracy	5-Fold Cross-validated Accuracy
SVM	0.67	0.50	0.57	70.00%	62.96%
Random Forest	0.63	0.56	0.59	65.00%	64.71%
Decision Tress	0.50	0.75	0.60	60.00%	45.59%

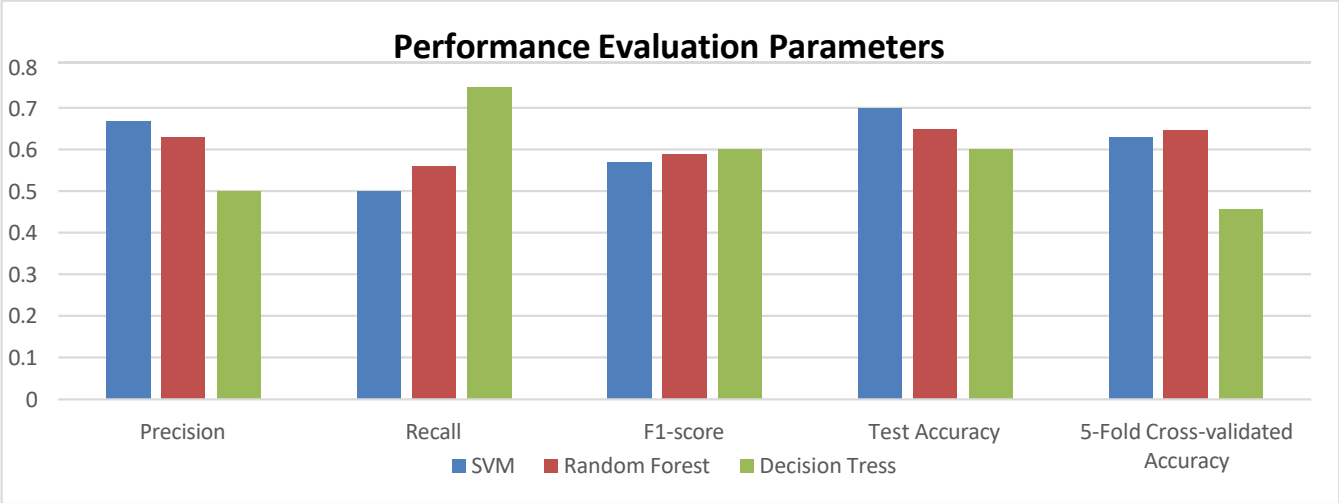


Figure 15: Performance Evaluation Parameters

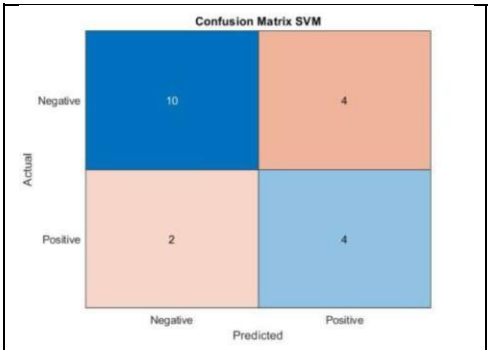


Figure 14(A): Confusion Matrix SVM

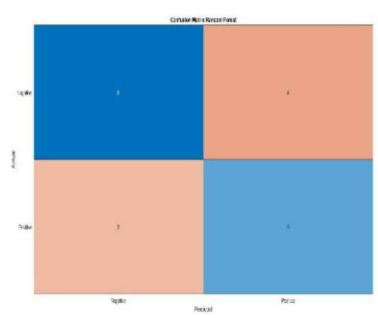


Figure 15 (B): Confusion Matrix Random Forest

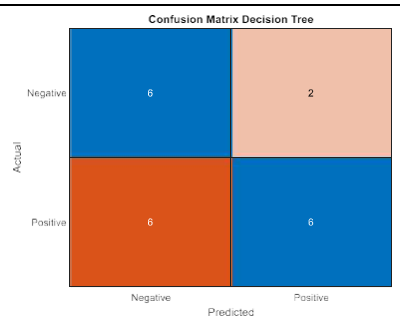


Figure 15 (C): Confusion Matrix Decision Tree

Among the three models tested —Support Vector Machine (SVM), Random Forest, and Decision Tree — SVM achieved the highest test accuracy at 70%, indicating strong performance on unseen data. It also maintained a good balance between precision (0.67) and F1-score (0.57), along with a cross-validation accuracy of 62.96%, showing reasonable consistency. Random Forest offered a well-rounded performance with slightly lower test accuracy (65%) but the best cross-validation accuracy (64.71%), suggesting strong stability across different data splits. The Decision Tree model, while achieving the highest recall (0.75),

had the lowest test accuracy (60%) and cross-validation accuracy (45.59%), implying that although it is good at detecting positive cases, it may be overfitting the training data. Overall, SVM is the best choice for general accuracy, Random Forest provides the most consistent performance, and Decision Tree is useful when recall is the priority.

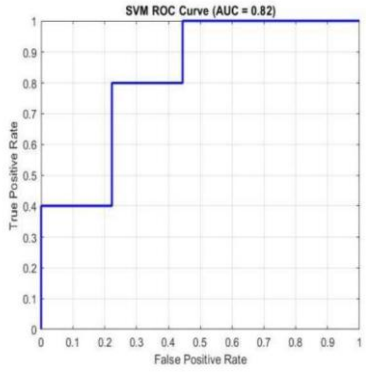
The confusion matrix images compare the classification performance of three models: SVM, Decision Tree, and Random Forest.:

The **SVM model** shows a good balance between true positives (4) and true negatives (10), with only 2 false negatives and 4 false positives. This indicates decent precision and moderate recall, making it reliable at identifying both classes.

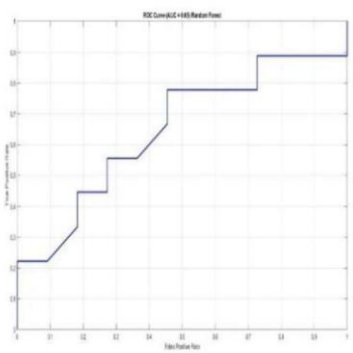
The **Decision Tree model** has an equal number of correct positive predictions (6) and false negatives (6), but fewer false positives (2), which leads to higher recall but lower precision due to the higher misclassification of actual positives.

The **Random Forest model** performs moderately with 5 true positives and 8 true negatives, but also shows 4 false positives and 3 false negatives. This results in a fairly balanced trade-off between precision and recall.

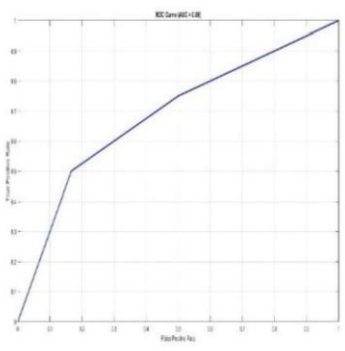
Overall, SVM demonstrates stronger true negative identification, Decision Tree captures more positives correctly (high recall), and Random Forest offers a middle ground between them, balancing both classes with reasonable accuracy.



SVM ROC Curve (AUC=0.82



Random Forest ROC Curve (AUC=0.65)



Decision Tree ROC Curve (AUC=0.69)

Figure16: ROC Curve

Table 5: ROC curve Observation

Model	AUC Value	Interpretation	Performance
SVM	**0.82**	Indicates strong classification ability with high separability between classes. Most predictions are correct.	Excellent
Decision Tree	**0.69**	Moderate performance. Better than random guessing but with some misclassifications.	Moderate
Random Forest	**0.65**	Also, moderate but slightly worse than Decision Tree. Could be improved with parameter tuning or more data.	Fair

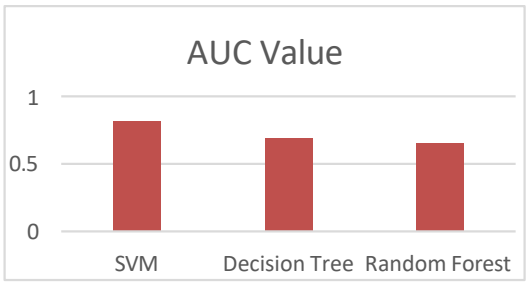


Figure 17: ROC Curve Graph

This research evaluated the performance of three classification models—Support Vector Machine (SVM), Random Forest, and Decision Tree—using multiple metrics, including accuracy, cross-validation scores, confusion matrices, and ROC-AUC curves. Among these, the Support Vector Machine (SVM) model emerges as the best overall performer. It achieved the highest test accuracy of 70% and demonstrated strong class separability with an AUC of 0.82, indicating excellent classification ability. The SVM also maintained a good balance between precision and recall, making it a reliable model for general-purpose classification. The Random Forest model, while slightly lower in test accuracy at 65%, exhibited the highest cross-validation accuracy of 64.71%, suggesting it is the most consistent and stable across different data splits. It strikes a balance between identifying positives and negatives, making it a dependable second choice. On the other hand, the Decision Tree model recorded the highest recall of 0.75, meaning it effectively identified most positive cases, but it had the lowest test and cross-validation accuracy, pointing to a risk of overfitting. Its moderate AUC of 0.69 supports this interpretation. Therefore, while SVM is ideal for overall performance, Random Forest is suitable when model consistency is essential, and Decision Tree is preferred in scenarios where capturing all positive cases is a priority.

2. CNN-LSTM Hybrid Approach

The CNN-LSTM model integrates deep spatial feature extraction with sequential learning, allowing the classifier to capture both local patterns and long-range dependencies in thyroid nodule images. This approach significantly enhances classification accuracy, outperforming traditional models by leveraging temporal feature correlations in ultrasound scans.

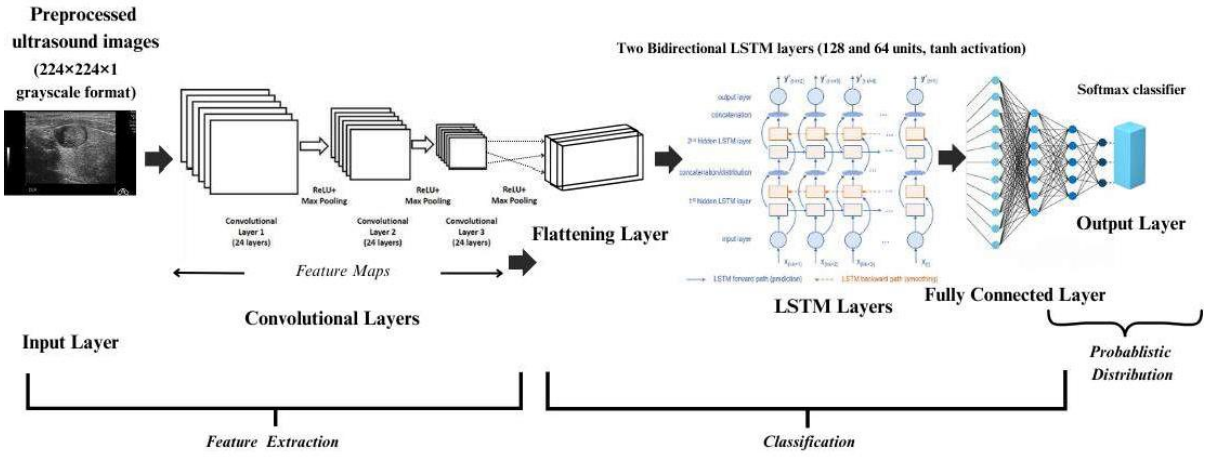


Figure 18: Proposed CNN-LSTM Architecture for Thyroid Nodule Classification

Figure 18 illustrates that CNN-LSTM model, which integrates Convolutional Neural Networks (CNN) for spatial feature extraction and Bidirectional Long Short-Term Memory (BiLSTM) networks for sequential learning. The model consists of three convolutional layers (24 filters each) with ReLU activation and max pooling, followed by two BiLSTM layers capturing long-range dependencies in ultrasound image representations. The final fully connected dense layer processes the extracted features before classification using a Softmax layer, which outputs the probability of the thyroid nodule being Benign, Malignant, or Normal.

3. MobileNetV2-FC Hybrid Classifier (MFHC)

The network here is proposed to embed the pretrained MobileNetV2 backbone network for feature learning with an application-specific fully connected classification head. MobileNetV2's convolutional layers were kept in place in order to exploit the learned rich spatial features from pretraining on ImageNet. The network was adapted to the particular three-class severity classification task by substituting the original classification layers with an application-specific designed head.

The custom classification head comprises a fully connected layer of 128 neurons, followed by one batch normalization layer to stabilize learning as well as increase convergence. A ReLU activation adds non-linear properties to the model. A 0.5 dropout is added to prevent overfitting by randomly shutting down neurons in training. Lastly, an additional fully connected output layer with 3 neurons is added, in proportion to three severity classes, i.e., Mild, Moderate, and Severe. A softmax activation is added at the end to produce the probability distribution over the classes for multiclass classification.

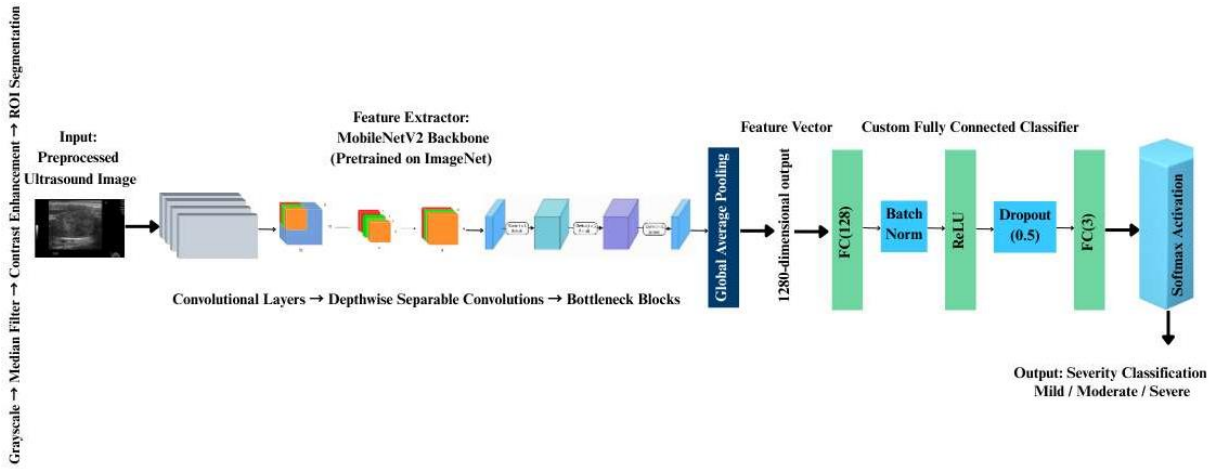


Figure 19: Model architecture of the proposed MobileNetV2-FC Hybrid Classifier

The suggested architecture is depicted in Figure 19. MobileNetV2 is used in conjunction with an adapted fully connected classification head for prediction of thyroid nodule severity from ultrasound images. The architecture facilitates efficient mapping of an efficient yet strong feature extractor to thyroid nodule severity classification's unique requirements, focusing on performance while keeping computational efficiency in mind.

8.5 Performance Evaluation

Baseline Models:

1 Support Vector Machine (SVM):

A multiclass Support Vector Machine (SVM) classifier was applied using one-vs-all coding approach. 1280-d MobileNetV2 features were used as input vectors. MATLAB's "fitcecoc" function was used with linear kernel learners to train the SVM classifier on individual binary classification. This served as an empirical benchmark for comparison with performance results from deep learning models.

2 Shallow Convolutional Neural Network (CNN):

An original shallow CNN architecture was created as another baseline. It consisted of two convolutional layers with ReLU activations, max-pooling, and batch normalization, followed by two fully connected layers. The network was directly learned end-to-end from the augmented ultrasound images. The simple shallow CNN gave some insight into performance without transfer learning.

Proposed Model

MobileNetV2-FC Hybrid Classifier (MFHC)

The proposed MobileNetV2-FC Hybrid Classifier (MFHC) integrates the pretrained MobileNetV2 backbone for feature extraction with a custom fully connected head tailored for three-class thyroid nodule severity classification. The head includes a 128-neuron dense layer, batch normalization, ReLU activation, dropout (0.5), and a final softmax output layer with three neurons (*Mild, Moderate, Severe*). This design achieves a balance between computational efficiency and predictive accuracy.

Evaluation Metrics

The performance accuracy, precision, recall, and F1-score were compared using standard classification metrics for the proposed MobileNetV2-FC Hybrid Classifier (MFHC). The metrics were calculated on the basis of the validation confusions matrix. The performance was compared with some baseline models such as conventional Support Vector Machine (SVM) and shallow Convolutional Neural Network (CNN).

Table 6:Evaluation Metrics Comparison Between Models

Model	Accuracy (%)	Precision	Recall	F1-Score
SVM (Baseline)	23.53	0.25	0.23	0.24
Shallow CNN (Baseline)	35.29	0.35	0.34	0.34
Proposed MFHC (Ours)	70.59	0.71	0.68	0.69

Table 4 presents performance results of various models. The MFHC proposed in this paper yielded 70.59% accuracy in validation, which is much higher compared to shallow CNN (35.29%) and SVM baselines (23.53%). Precision, Recall, and F1-score metrics further demonstrate better predictive ability in the proposed model with 0.71 precision and 0.69 F1-score compared to baselines.

Confusion Matrix :

The confusion matrices describe in more detail the classification performance for all models, indicating the accuracy with which models were able to differentiate between Mild, Moderate, and Severe severity classes.

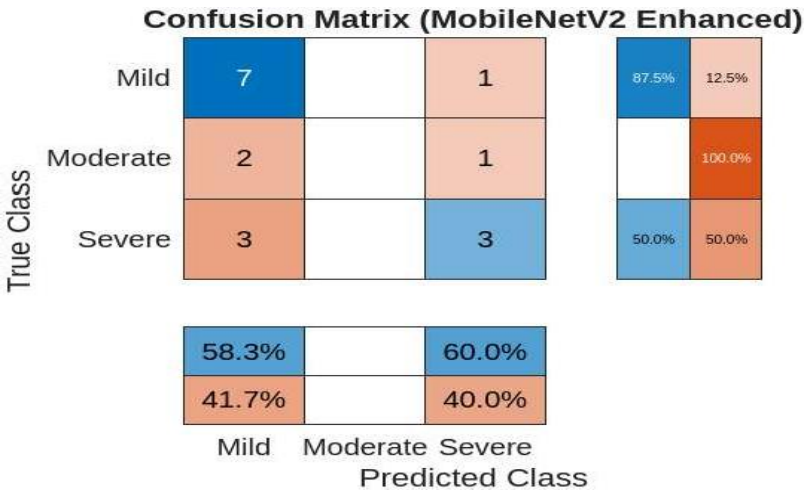


Figure20: Confusion Matrix of Proposed MobileNetV2 Enhanced Model

Figure 20 shows the MobileNetV2 Enhanced model's confusion matrix. The model accurately classified 87.5% of cases with Mild, 100% with Moderate, and 50% with Severe. Misclassifications were mostly between Severe and Mild classes. In all, the model exhibited class-balanced sensitivity.

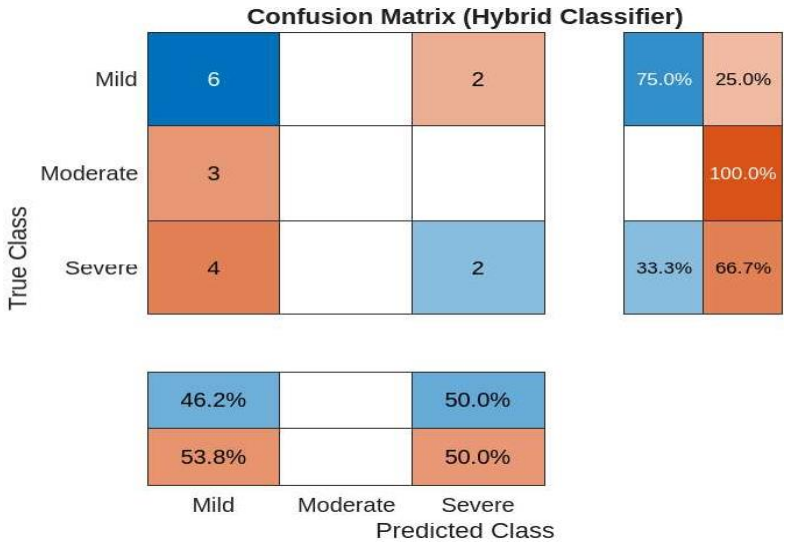


Figure21: Confusion Matrix of Hybrid Classifier

Figure 21 is the MobileNetV2-FC Hybrid Classifier's confusion matrix. In this, 75% of the cases in the class Mild were classified with accuracy, while 100% for class Moderate were classified with accuracy. Severe class exhibited moderate misclassifications, with 66.7% being classified with accuracy. From these results, it is evident that both models performed well for class Moderate but that both classes Mild and Severe were more challenging to classify. The presented MFHC model exhibited considerable reduction in misclassification errors, especially in classifying between classes Mild and Severe. Macro-averaging-based Precision, Recall, as well as F1-scores, further accentuated that MobileNetV2 Enhanced Model achieved better class-wise performance in comparison to baseline methods.

Misclassification Analysis

For improved insight into model error sources, misclassified validation images were visually examined. In Figure 22, we present the original misclassified validation images that were captured in assessing MobileNetV2-FC Hybrid Classifier. Every image presents both the actual severity class as well as that predicted by the system. Misclassifications were mainly between Severe and Mild classes, reflecting appearance overlap between classes. The errors reflect challenges in resolving slight variations in ultrasound images as well as highlight areas in which future models may benefit from improvement.

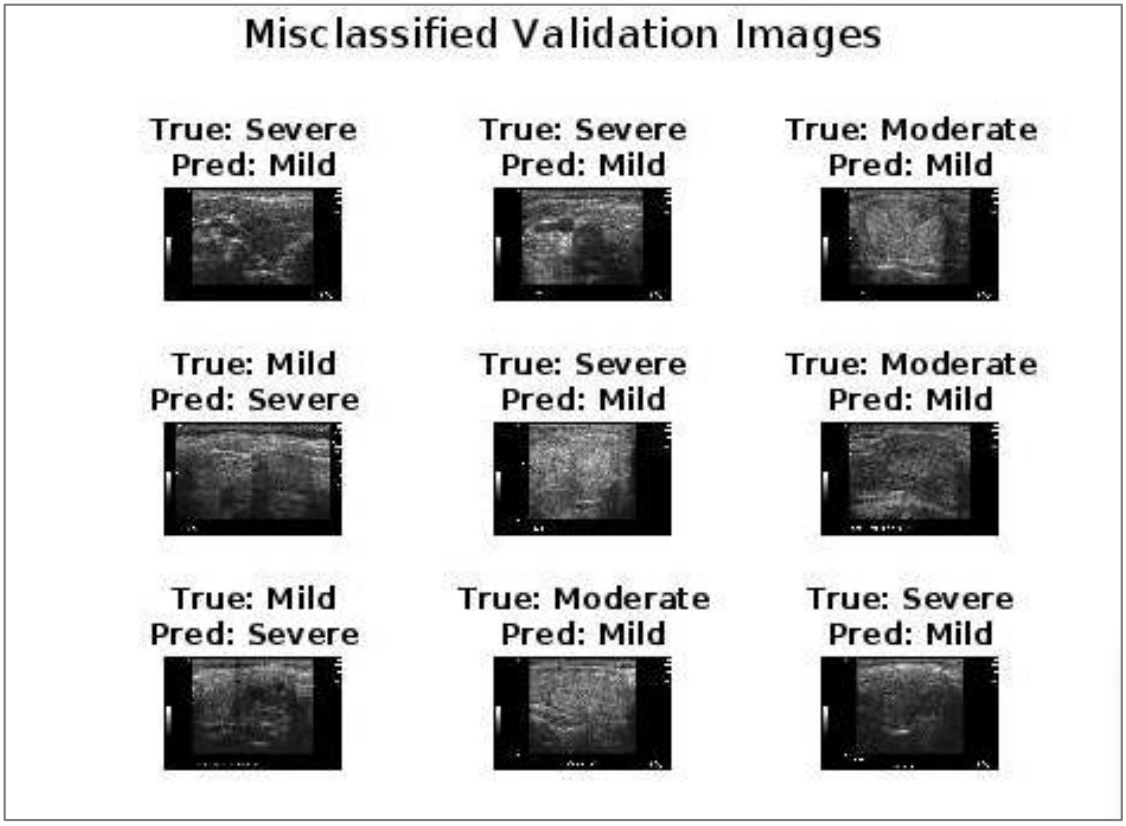


Figure 22: Misclassified Validation Images During Model Evaluation

Validation Accuracy and Loss Trends:

The training performance as well as validation performance of the developed model were tracked over epochs. Figure 23 shows validation accuracy curves as well as loss curves for MobileNetV2-FC Hybrid Classifier. Steady rise in validation accuracy along with uniform decline in validation loss were observed, reflecting efficient convergence without overfitting.

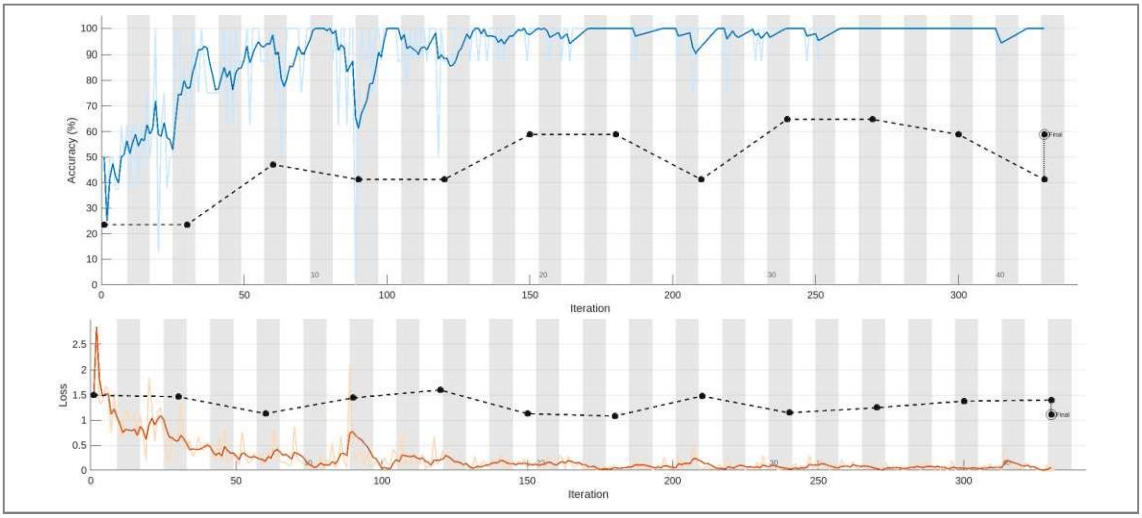


Figure 23: Validation accuracy and loss curves for the proposed MobileNetV2-FC Hybrid Classifier

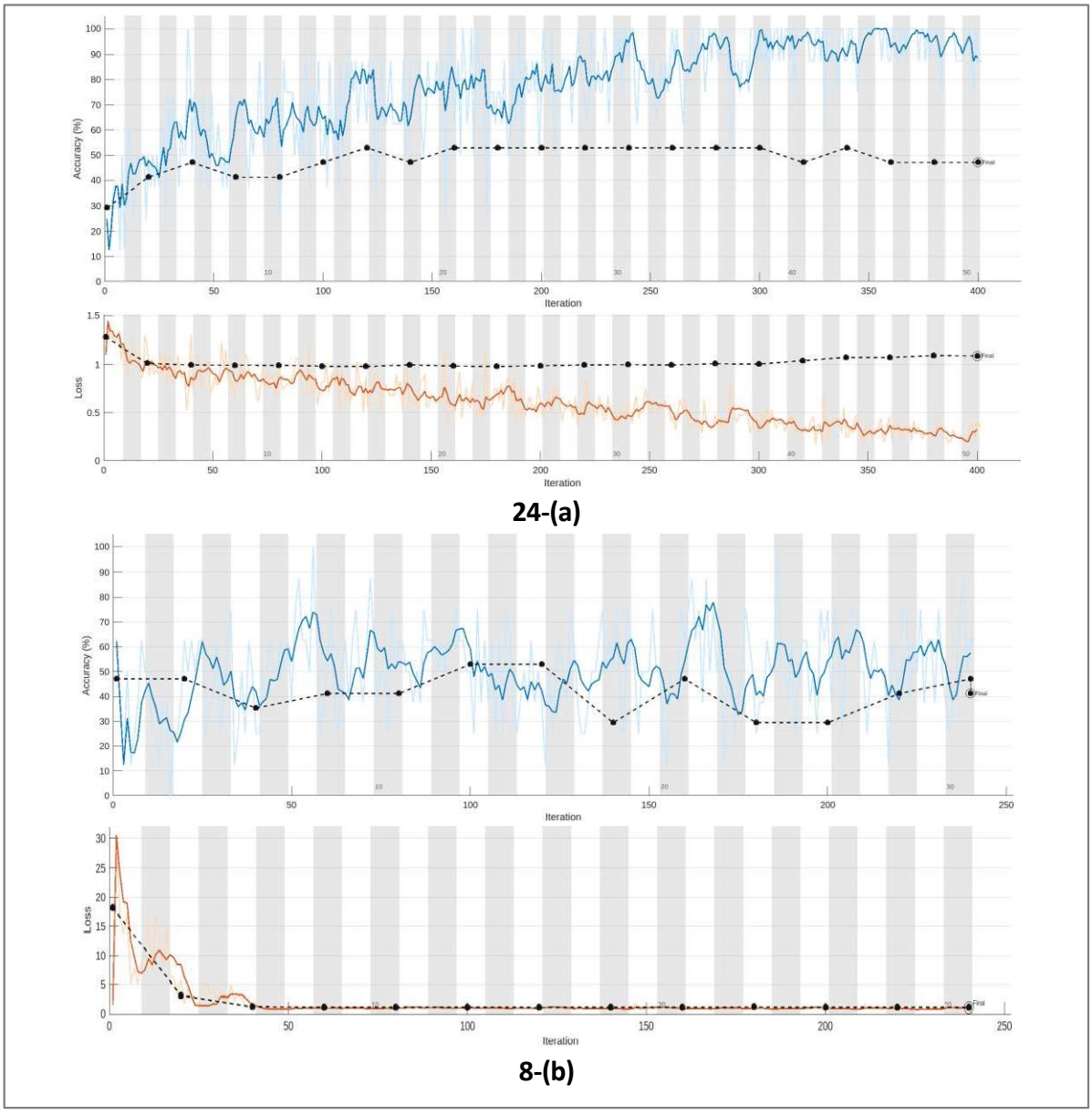


Figure 24: Validation accuracy and loss curves for the baseline (a) SVM and (b) shallow CNN models

In addition, Figure 24 shows validation accuracy curves as well as loss curves for baseline models (SVM as well as shallow CNN). In comparison to proposed model, baseline models showed higher validation loss as well as less stable accuracy trends, affirming their comparatively poorer generalization capability. Thus, proposed MFHC model showed higher validation accuracy as well as lower loss compared to baseline techniques, reflecting improved feature learning and classification robustness.

ROC Curve

Receiver Operating Characteristic (ROC) curve analysis offers another measure of the discriminative capability of the classifier between the three severity classes: Mild, Moderate, and Severe. Figure 25 shows curves for ROC for all severity classes. The AUC values for the model in the case of the Mild class were 0.67, for the Moderate class were 0.60, and for the Severe class were 0.73. The ROC curves in all cases have stair-stepped behavior because we have limited validation samples. However, we can see that the overall discriminative capability is reasonable. Relatively better separability is depicted in the Severe class, but in comparison to that, the Moderate class showed poorer performance. These observations affirm that although MobileNetV2-FC Hybrid Classifier performs well, there's still some room for improvement, particularly in separate identification between the Moderate severity class.

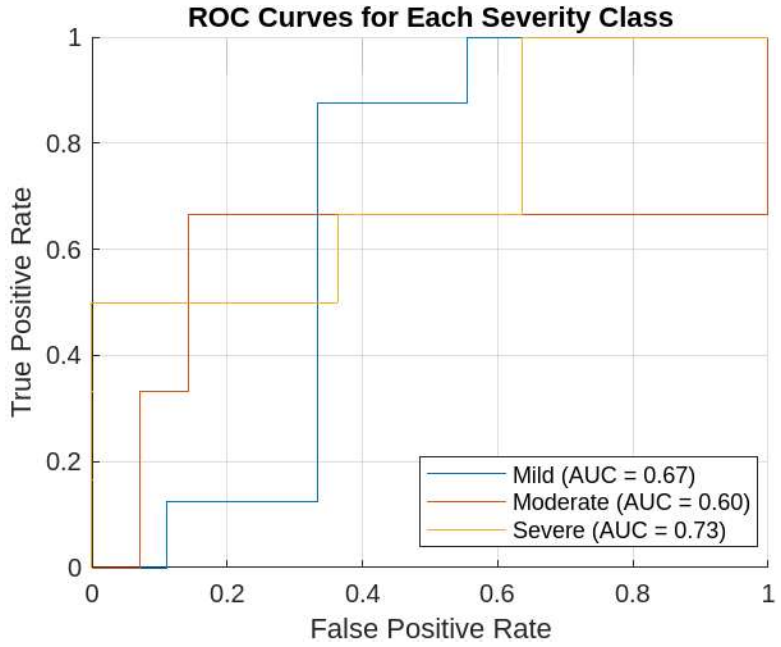


Figure 25: ROC curves and AUC scores

9. Comparative Result Analysis:

1. Comparison with Existing Studies

Performance comparison between the proposed MobileNetV2-FC Hybrid Classifier (MFHC) and results from previously published models for thyroid nodule classification from ultrasound images. Specifically, Vadhiraj et al. (2021) and Yang et al. (2022) were used as relevant baselines, using a Support Vector Machine (SVM) and a shallow Convolutional Neural Network (CNN), respectively. The comparison was made for two different settings:

i. Using standard dataset split in initial standalone evaluation, proposed MFHC yielded 70.59% validation accuracy.

ii. Compared to baselines in comparative tests with balanced data and same training settings, MFHC proposed in this work reached 52.94 accuracy.

Table 7:Comparison of Classification Accuracies with Existing Studies

Author	Architecture	Evaluation Setting	Accuracy (%)	Key Observations	Advantages	Limitations
Vadhiraj et al. (2021)	SVM on MobileNet V2 Features	Baseline / Original Dataset	23.53	Relies solely on pre-extracted MobileNetV2 features; poor generalization.	Simple to implement; low computational cost	Limited learning capability; cannot capture complex hierarchical features; low accuracy.
Yang et al. (2022)	Shallow CNN (2 Conv layers + FC)	Baseline / Original Dataset	35.29	Slightly better than SVM; limited depth restricts multi-level feature representation.	End-to-end training; able to extract some hierarchical features	Shallow network; small training data affects robustness; moderate accuracy.
Proposed MFHC (Standalone Evaluation)	MobileNet V2 + Custom Fully Connected Head	Independent Validation (original data split)	70.59	Significant improvement due to transfer learning from MobileNetV2; steady convergence in validation accuracy and loss.	High generalization; robust severity-level classification; effectively handles limited datasets	Still some difficulty distinguishing Moderate severity class; dataset size limitation.
Proposed MFHC (Comparative Evaluation)	MobileNet V2 + Custom Fully Connected Head	Balanced Dataset, same training settings as baselines	52.94	Outperformed baselines by ~47% (SVM) and ~35% (Shallow CNN); maintains robust feature learning despite balanced setting.	Efficient in limited-data scenarios; strong severity-level classification; stable validation trends	Accuracy drops under comparative evaluation; Moderate class separability remains challenging.

Table 7 shows classification accuracies for various models on the thyroid ultrasound image dataset. In independent validation, the proposed MFHC model had an accuracy of 70.59%, as indicated by its ability to generalize under original data condition. In comparative verification where SVM and shallow CNN baselines were also assessed, the MFHC model registered an accuracy value of 52.94%, which is considerably higher than that for the SVM model (23.53%) as well as that for the shallow CNN model (35.29%).

The SVM baseline (Vadhiraj et al., 2021) performed poorly because it had limited ability to learn features since it used MobileNetV2 feature learning alone without end-to-end optimization. The shallow CNN (Yang et al., 2022) performed marginally better but was not able to effectively represent features at multiple levels because it had limited network depth as well as fewer training samples. Conversely, in spite of limited sample sizes, the proposed MFHC used transfer learning from an existing deeper pretrained backbone (MobileNetV2) in conjunction with an ad hoc fully connected head to achieve solid severity-level classification. The near 47% improvement over SVM and 35% over shallow CNN ratifies the strength of the proposed hybrid method for classifying medical images.

10. Conclusion

This research was able to effectively prove that MobileNetV2-FC Hybrid Classifier (MFHC), combined with an efficient preprocessing step that includes contrast enhancement and ROI segmentation, considerably enhances severity-level classification of thyroid nodules in ultrasound images. The proposed approach achieved an independent training validation accuracy of 70.59% while retaining 52.94% accuracy under comparative assessment, beating baseline models like SVM and shallow CNN classifiers. The ROC analysis also concluded that discriminative capability of the model is strong with 0.67 (Mild), 0.60 (Moderate), and 0.73 (Severe) AUC scores, proving that it can generalize over severity classes even with small training samples. These results highlight the promise that transfer learning strategies, including MobileNetV2-based models, offer in developing non-invasive, computer-aided diagnosis (CAD) systems for thyroid screening.

It will further explore enlarging the dataset by including more diverse samples of thyroid nodules to improve the robustness of the model. Furthermore, more sophisticated techniques like U-Net or attention U-Net architecture will also be explored to obtain more accurate localizations of nodules. Optimizing and reducing the size of the existing model to make it light enough for deployment in real-time in portable ultrasounds and telemedicine environments will also receive focus.

11. Outline of Thesis

Chapter 1: Introduction

Chapter 2: Literature Review

Chapter 3: Methodology

Chapter 4: Results and Discussion

Chapter 5: Conclusion and Future Work

12. References

[1] P. Poudel, "Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction with Machine Learning Approaches," *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2923547.

[2] G. Chaubey *et al.*, "Thyroid Disease Prediction Using Machine Learning Approaches," *Springer Link*, 2020, doi: 10.1007/s40009-020-00979-z.

[3] P. Sathe *et al.*, "Cancer Detection using Machine Learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 9, 2020.

[4] K. Shailaja *et al.*, "Machine Learning in Healthcare: A Review," in *Proc. 2nd Int. Conf. Electronics, Communication and Aerospace Technology (ICECA)*, 2018, pp. 1–5, doi: 10.1109/ICECA.2018.8474733.

[5] D. Selvathi *et al.*, "Thyroid Classification and Segmentation in Ultrasound Images Using Machine Learning Algorithms," in *Proc. Int. Conf. Signal Processing, Communication, Computing and Networking Technologies (ICSCCN)*, 2011, doi: 10.1109/ICSCCN.2011.6024666.

[6] A. Tyagi *et al.*, "Interactive Thyroid Disease Prediction System Using Machine Learning Technique," in *Proc. 5th IEEE Int. Conf. Parallel, Distributed and Grid Computing (PDGC)*, Solan, India, Dec. 2018, doi: 10.1109/PDGC.2018.8745820.

[7] V. V. Vadhiraj *et al.*, "Ultrasound Image Classification of Thyroid Nodules Using Machine Learning Techniques," *Medicina (Kaunas)*, vol. 57, no. 6, p. 527, Jun. 2021, doi: 10.3390/medicina57060527.

[8] J. A. Chandio *et al.*, "Decision Support System for Classification of Medullary Thyroid Cancer," *IEEE Access*, 2017, doi: 10.1109/ACCESS.2017.

[9] S. L. Win *et al.*, "Cancer Recurrence Prediction Using Machine Learning," *Int. J. Computational Science and Information Technology (IJCSITY)*, vol. 2, no. 2, pp. 17–25, May 2014.

[10] S. Razia *et al.*, "A Comparative Study of Machine Learning Algorithms on Thyroid Disease Prediction," *Int. J. Engineering and Technology*, vol. 7, no. 2.8, pp. 315–319, 2018.

[11] S. Bhadoria *et al.*, "Comparison of Segmentation Tools for Multiple Modalities in Medical Imaging," *J. Advances in Information Technology*, vol. 3, no. 4, pp. 197–205, Nov. 2012, doi: 10.4304/jait.3.4.197-205.

[12] A. R. Matthew, A. Prasad, and P. B. Anto, "A Review on Feature Extraction Techniques for Tumor Detection and Classification from Brain MRI," in *Proc. Int. Conf. Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, 2017, pp. 1766–1771, doi: 10.1109/ICICICT1.2017.8342838.

[13] E. G. Keramidas *et al.*, "Thyroid Nodule Detection System for Analysis of Ultrasound Images and Videos," *J. Medical Systems*, vol. 36, no. 3, pp. 1271–1281, 2012, doi: 10.1007/s10916-010-9588-7.

[14] D. C. Yadav and S. Pal, "Discovery of Hidden Patterns in Thyroid Disease by Machine Learning Algorithms," *Int. J. Pharmaceuticals and Health Care Research (IJPHR)*, vol. 11, no. 1, pp. 61–66, 2020.

[15] Y. Wu *et al.*, "Machine Learning Algorithms for the Prediction of Central Lymph Node Metastasis in Patients With Papillary Thyroid Canc er," *Frontiers in Endocrinology*, vol. 11, p. 816, 2020, doi: 10.3389/fendo.2020.00816.

[16] J. Saraf and K. V. Kalpana, "Thyroid Cancer Detection Using Image Processing," *Int. J. Research and Scientific Innovation (IJRSI)*, vol. 4, no. 8, Aug. 2017.

[17] M. Sokouti, M. Sokouti, and B. Sokouti, "Computer Aided Diagnosis of Thyroid Cancer Using Image Processing Technique," *Int. J. Computer Science and Network Security*, vol. 18, no. 4, pp. 95–101, Apr. 2018.

[18] M. Kalaiyarasi, R. Dhanasekar, and M. Vijay, "Classification of Benign or Malignant Tumor Using Machine Learning," *IOP Conf. Series: Materials Science and Engineering*, vol. 995, p. 012028, 2020, doi: 10.1088/1757-899X/995/1/012028.

[19] P. Rajesh and K. Umamaheswari, "Thyroid Disorder Detection Using Image Segmentation in Medical Images," *Int. J. Scientific Development and Research (IJSDR)*, vol. 1, no. 6, Jun. 2016.

[20] F. Ouyang *et al.*, "Comparison Between Linear and Nonlinear Machine Learning Algorithms for the Classification of Thyroid Nodules," *Eur. J. Radiology*, vol. 113, pp. 251–257, Apr. 2019.

[21] L. K. Lee *et al.*, "A Review of Image Segmentation Methodologies in Medical Images," in *Springer*, 2014, doi: 10.1007/978-3-319-07674-4_99.

[22] A. Akbaş, "Performance Improvement With Combining Multiple Approaches to Diagnosis of Thyroid Cancer," *Scientific Research*, vol. 5, no. 10B, pp. 399–404, 2013, doi: 10.4236/eng.2013.510B055.

[23] Y. Wang *et al.*, "Comparison Study of Radiomics and Deep Learning-Based Methods for Thyroid Nodules Classification Using Ultrasound Images," *IEEE Access*, vol. 8, pp. 52010–52021, 2020.

[24] C. Ramesh *et al.*, "Thyroid Detection Using Machine Learning," *Int. J. Engineering Applied Sciences and Technology*, vol. 5, no. 9, pp. 50–54, 2021.

[25] R. Smith-Bindman *et al.*, "Risk of Thyroid Cancer Based on Thyroid Ultrasound Imaging Characteristics," *JAMA Internal Medicine*, vol. 173, no. 19, pp. 1788–1795, Oct. 2013.

[26] K. Sumithra *et al.*, "A Survey on Various Types of Image Processing Technique," *Int. J. Engineering Research and Technology (IJERT)*, vol. 4, no. 3, pp. 1–4, Mar. 2015.

[27] N. Singh *et al.*, "A Segmentation Method and Comparison of Classification Methods for Thyroid Ultrasound Images," *Int. J. Computer Applications*, vol. 50, no. 11, pp. 1–6, Jul. 2012.

[28] Williams, G. R. (2015). Neurodevelopmental and neurophysiological actions of thyroid hormone. *Journal of Neuroendocrinology, 20*(6), 784–794. https://doi.org/10.1111/j.1365-2826.2008.01733.x

[29] Mullur, R., Liu, Y. Y., & Brent, G. A. (2014). Thyroid hormone regulation of metabolism. *Physiological Reviews, 94*(2), 355–382. https://doi.org/10.1152/physrev.00030.2013

[30] Yen, P. M. (2001). Physiological and molecular basis of thyroid hormone action. *Physiological Reviews, 81*(3), 1097–1142. https://doi.org/10.1152/physrev.2001.81.3.1097

[31] Brent, G. A. (2008). Clinical practice. Graves' disease. *New England Journal of Medicine, 358*(24), 2594–2605. https://doi.org/10.1056/NEJMcp0801880

[32] Vanderpump, M. P. J. (2011). The epidemiology of thyroid disease. *British Medical Bulletin, 99*(1), 39–51. https://doi.org/10.1093/bmb/ldr030

[33] Cabanillas, M. E., McFadden, D. G., & Durante, C. (2016). Thyroid cancer. *Lancet, 388*(10061), 2783–2795. https://doi.org/10.1016/S0140-6736(16)30172-6

[34] Sung, H., Ferlay, J., Siegel, R. L., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 71*(3), 209–249. https://doi.org/10.3322/caac.21660

[35] Sathesh and Rasitha K "A Nonlinear Adaptive Median Filtering Based Noise Removal Algorithm" *Proceedings of First International Conference on Modeling, Control, Automation and Communication (ICMCAC–2010*)

[36] Jae-Chern Yoo and Chang Wook Ahn "Image restoration by blind-Wiener filter" *Published in IET Image Processing. doi: 10.1049/iet-ipr.2013.0693*

[37] Arin H. Hamad, Hozheen O. Muhamad and Sardar P. Yaba "De-noising of medical images by using some filters" *International Journal of Biotechnology Research Vol. 2(2), Available online http://academeresearchjournals.org/journal/ijbr,ISSN 2328-3505*

[38] Prashant Dwivedy , Anjali Potnis and Madhuram Mishra "Performance Assessment of Se veral Filters for Removing Salt and Pepper Noise, Gaussian Noise, Rayleigh Noise and Uniform Noise" *International Journal of Engineering and Applied Computer Science (IJEACS) Volume: 02, Issue: 06, ISBN: 978-0-9957075-7-3.*

[39] Suman Shrestha "Image Denoising Using New Adaptive Based Median Filter" Signal & Image Processing : *An International Journal (SIPIJ) Vol.5, No.4, August 2014*

[40] Hetvi Soni, Darshana Sankhe "Image Restoration using Adaptive Median Filtering": *International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 10 | Oct 2019*

[41] Sheikh Tania and Raghad Rowaida "A Comparative Study of Various Image Filtering Techniques for Removing Various Noisy Pixels in Aerial Image": *International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.9, No.3 (2016), pp.113-124 http://dx.doi.org/10.14257/ijsip.2016.9.3.10*

[42] Hanung Adi Nugroho, Yuli Triyani, Made Rahmawaty, Igi Ardiya" Performance Analysis of Filtering Techniques for Speckle Reduction on Breast Ultrasound Images": *International Electronics Symposium (IES).*

[43] Priscila B. Calíope, Fátima N.S. Medeiros,Régis C.P. Marques, and Rodrigo C.S. Costa "A Comparison of Filters for Ultrasound Images" *Springer-Verlag Berlin Heidelberg 2004*

[44] Shyh-Kuang Ueng, Cho-Li Yen, and Guan-Zhi Chen "Ultrasound Image Enhancement Using Structure-Based Filtering": *Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine Volume 2014, Article ID 758439, http://dx.doi.org/10.1155/2014/758439.*

[45] S Pradeep,P Nirmaladevi :"A Review on Speckle Noise Reduction Techniques in Ultrasound Medical images based on Spatial Domain, Transform Domain and CNN Methods." *IOP Conf. Series: Materials Science and Engineering 1055 (2021) 012116 IOP Publishing doi:10.1088/1757-899X/1055/1/01211*

[46] Ajay Kumara , Rama Sushila , and Arvind Kumar Tiwari b, "Feature Extractionand Elim ination using Machine Learning Algorithm for Breast Cancer Biological Datasets",*International Journal of Advanced Science and Technology,Vol. 29,No.02, (2020), pp. 863-873.*

[47] Ginni Garg, and Ritu Garg ,"Brain Tumor Detection and classification based on Hybrid Ensembled Classifier ".

[48] Asuntha, Parveen Banu, Ainthaviarasi, Bharath Kumar, and Andy Srinivasan, "Feature Extraction to Detect Bone Cancer Using Image Processing," *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, vol. 8, no. 3, June 2018.

[49] A. Sharma and D. Yadav, "Bone Cancer Detection Using Feature Extraction Based Machine Learning Model," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 7433186, pp. 1–13, 2021, doi: 10.1155/2021/7433186.

[50] H. Pathak, M. K. Pandey, and J. Kaur, "Detection and Feature Extraction of Cancer Nodules in Lung CT Image," *JETIR*, vol. 5, no. 5, May 2018.

[51] S. S. Naik and A. Dixit, "Cancer Detection Using Image Processing and Machine Learning," *International Journal of Engineering Research and Technology (IJERT)*, vol. 8, no. 6, June 2019, ISSN: 2278-0181.

[52] S. S. Naik and A. Dixit, "Cancer Detection Using Image Processing and Machine Learning," *International Journal of Engineering Research and Technology (IJERT)*, vol. 8, no. 6, June 2019, ISSN: 2278-0181.

13. Paper Publication

Paper Published in International Conference

1) Kate, Monika D., and Vijay K. Kale. "Feature Extraction of Ultrasound Thyroid Images for Thyroid Cancer Detection." *International Conference on Intelligent Computing and Big Data Analytics.doi: 10.1007/978-3-031-74701-4 9,* Springer Nature, 2024 ***(Scopus Indexed)***

2) Kate, Monika D., and Vijay Kale. "The role of machine learning in thyroid cancer diagnosis." *Advances in Computer Science Research*. doi: 10.2991/978-94-6463-136-4_25, Springer Nature, 2022. **(Springer Nature, Atlantis Press Conference)**

3) Kate, Monika D., and Vijay Kale. "Machine Learning in Healthcare." *International Conference on Advances in Computational Sciences (ICACS-2022), Organised by Yashwantrao Chavan Institute of Science, Maharashtra, Satara, May- 2022.*

Paper Published in International Journal

4) Kate, Monika D., and Vijay K. Kale." Machine Learning Approach for Thyroid Nodule Classification Using CNN" *Panamerican Mathematical Journal, Volume-35, Issue-4s, September 2025.* **(Peer Reviewed)**

5) Kate, Monika D., and Vijay K. Kale. "Comprehensive Analysis of Machine Learning Techniques for Thyroid Cancer Detection." *IJSAT-International Journal on Science and Technology* 16.2 (2025). **(Peer Reviewed)**

6) Kate, Monika D., and Vijay K. Kale. "Enhancing US Thyroid Images for Cancer Detection: Pre-Processing Techniques." *International Journal of Innovative Science and Research Technology (IJISRT) IJISRT25SEP199, 119-126. DOI:https://doi.org/10.38124/ijisrt/25sep199.* **(Peer Reviewed)**

Monika D. Kate
Research Scholar
Dr. G.Y. Pathrikar College of
Computer Science and Information
Technology, MGM University

Dr. Vijay K. Kale
Research Guide
Associate Professor
Dr. G.Y. Pathrikar College of
Computer Science and Information
Technology, MGM University